DOCUMENT RESUME

ED 173 435 TM 009 572

TITLE	Proceedings of the Invitational Conference on Testing Problems (10th, New York, New York, October 30, 1954).
INSTITUTION	Educational Testing Service, Princeton, N.J.
PUB DATE	30 Oct 54
NOTE	130p.
EDRS PRICE	MF01/PC06 Plus Postage.
DESCRIPTORS	*Academically Gifted; *Accelerated Programs; Advanced Placement Programs; Analysis of Variance; Decision Making; Educational Testing; Elementary Secondary Education; Factor Analysis; Group Dynamics; *Group Relations; Higher Education; Individual Differences; *Information Theory; Job Placement; Manpower Utilization; Predictor Variables; Psychological Testing; *Psychometrics; Social Relations; *Testing Problems
IDENTIFIERS	Latent Structure Analysis; Russell Sage Social Relations Test

ABSTRACT

Four major topics were discussed at the conference: application of information theory to testing; recent advances in psychometric methods; evaluating group interaction; and new developments in the education of abler students. Papers delivered were: Multiple Assignment of Persons to Jobs, by Paul S. Dwyer; New Light on Test Strategy from Decision Theory, by Lee J. Cronbach; The Relation Between Uncertainty and Variance, by William J. McGill; Some Recent Results in Latent Structure Analysis, by T. W. Anderson; The "Moderator Variable" as a Useful Tool in Prediction, by David R. Saunders; A Method of Factoring Without Communalities, by L. L. Thurstone; A New Technique for Measuring Individual Differences in Conformity to Group Judgment, by Richard S. Crutchfield; The Russell Sage Social Relations Test: A Measure of Group Problem-Solving Skills in Elementary School Children, by Dora E. Damrin; Description of Group Characteristics, by John K. Hemphill; Acceleration: Basic Principles and Recent Research, by Sidney L. Pressey; College Admission with Advanced Standing, by William H. Cornog; and Special Treatment for Abler Students and Its Relation to National Manpower, by Dael L. Wolfle. (MH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U S DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED173435

1954

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Educational
Testing Service

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM009 572

EDUCATIONAL TESTING SERVICE

BOARD OF TRUSTEES, 1954-1955

Lewis W. Jones, *Chairman*

Arthur S. Adams
Samuel T. Arnold
Frank H. Bowles
Doak S. Campbell
Charles W. Cole
Donald K. David
Clark Kerr
Katharine E. McBride
Thomas R. McConnell
Edward S. Noyes
William G. Saltonstall
George D. Stoddard

Benjamin C. Willis

OFFICERS

Henry Chauncey, *President*
Richard H. Sullivan, *Vice President and Treasurer*
William W. Turnbull, *Vice President*
Henry Dyer, *Vice President*
Jack K. Rimalover, *Secretary*
Catherine G. Sharp, *Assistant Secretary*
Robert F. Kolkebeck, *Assistant Treasurer*

COPYRIGHT, 1955, EDUCATIONAL TESTING SERVICE
20 NASSAU STREET, PRINCETON, N. J.
PRINTED IN THE UNITED STATES OF AMERICA
Library of Congress Catalog Number: 47-11220

INVITATIONAL CONFERENCE
ON
TESTING PROBLEMS

OCTOBER 30, 1954

EDWARD E. CURETON, *Chairman*

Application of Information Theory to Testing
Recent Advances in Psychometric Methods
Evaluating Group Interaction
New Developments in the Education of Abler Students

EDUCATIONAL TESTING SERVICE
PRINCETON, NEW JERSEY LOS ANGELES, CALIFORNIA

FOREWORD

The 1954 Invitational Conference on Testing Problems, the tenth of these meetings and the seventh sponsored by Educational Testing Service, was one of the most successful. Speakers and invited guests from near and far participated in a program of wide scope and significance. This published record of the proceedings will, we hope, carry the words and wisdom of the participants to an even greater audience.

To Edward E. Cureton, Chairman of the 1954 Conference, goes full credit for a job well done in spite of an arduous professional schedule. His imagination, energy, and attention to detail were primarily responsible for the well-planned and well-conducted program. His successful efforts are deeply appreciated by the many who enjoyed the 1954 Conference.

Henry Chauncey
President

PREFACE

The 1954 Invitational Conference on Testing Problems, sponsored by Educational Testing Service, was held in New York City at the Roosevelt Hotel as in past years, on October 30. This volume provides a permanent record of the papers and discussions.

In arranging the Conference program, it was necessary to try to balance two pairs of conflicting considerations. In recent years the attendance has risen almost to convention proportions. More than 1000 invitations are sent out, and almost half of those invited are usually present. The first problem, then, was to try to preserve something of the original flavor of an invitational conference, and at the same time try to arrange a program which would appeal to some 500-odd people of diverse interests and backgrounds.

At the technical level, measurement theory is undergoing rapid revisions under the impact of communication theory, information theory, statistical decision theory, latent structure analysis, factor analysis, and scale theory, to name only a few. These theories are all highly mathematical, and the branches of mathematics which they use are not well known to most workers in educational and psychological measurement. In the broader field of evaluation, moreover, clinical psychologists and social psychologists are developing new methods of assessing such things as creative talents, personality traits, the dimensions of group interaction, the nature and quality of leadership in various settings, and the processes of human judgment. These methods are in many cases quite different from those employed in assessing cognitive aptitudes and school achievement. It was felt that the new theories and methods should be brought to the attention of measurement workers despite the rather considerable obstacles to effective communication. The second problem was to try to reduce these obstacles as much as possible, and to find speakers who could present some of the new theories and methods in terms which measurement workers could understand.

We were fortunate in securing as our luncheon speaker Dr. Daniel Starch, whose address, ". . . And Have Not Wisdom," recalled forcibly the need to teach students how to make ethical value judgments, and the need to develop methods for measuring the attainment of this vital educational objective.

The rest of the program consisted of a first morning session on some applications of information theory to testing problems, two parallel sessions later in the morning, one on recent advances in psychometric methods and one on the evaluation of group interaction, and an afternoon session on new developments in the education of abler students. We hope that this program achieved the balances implied by its objectives.

The Chairman welcomes this opportunity to express his sincere appreciation to all the speakers for their contributions, to Educational Testing Service for sponsoring the Conference, to Jack K. Rimalover for his unfailing support, assistance, and counsel, and to Mrs. Catherine G. Sharp for her assistance in making such excellent local arrangements.

Edward E. Cureton, *Chairman*
1954 Conference

CONTENTS

FOREWORD by Henry Chauncey 3

PREFACE by Edward E. Cureton 5

GENERAL MEETING

"Application of Information Theory to Testing"

MULTIPLE ASSIGNMENTS OF PERSONS TO JOBS 11
Paul S. Dwyer, *University of Michigan*

NEW LIGHT ON TEST STRATEGY FROM DECISION THEORY 30
Lee J. Cronbach, *University of Illinois*

THE RELATION BETWEEN UNCERTAINTY AND VARIANCE 37
William J. McGill, *Massachusetts Institute of Technology*

SECTIONAL MEETINGS

SECTION I: *"Recent Advances in Psychometric Methods"*
Chairman and Discussion Leader:
Phillip J. Rulon, *Harvard University*

SOME RECENT RESULTS IN LATENT STRUCTURE ANALYSIS 49
T. W. Anderson, *Columbia University*

THE "MODERATOR VARIABLE" AS A USEFUL TOOL IN PREDICTION 54
David R. Saunders, *Educational Testing Service*

A METHOD OF FACTORING WITHOUT COMMUNALITIES 59
L. L. Thurstone, *University of North Carolina*

SECTION II: *"Evaluating Group Interaction"*
Chairman and Discussion Leader:
Irving Lorge, *Columbia University*

A NEW TECHNIQUE FOR MEASURING INDIVIDUAL DIFFERENCES IN CONFORMITY TO GROUP JUDGMENT 69
Richard S. Crutchfield, *University of California*

THE RUSSELL SAGE SOCIAL RELATIONS TEST: A MEASURE OF GROUP PROBLEM-SOLVING SKILLS IN ELEMENTARY SCHOOL CHILDREN 75
Dora E. Damrin, *Educational Testing Service*

DESCRIPTION OF GROUP CHARACTERISTICS 85
John K. Hemphill, *Ohio State University*

LUNCHEON ADDRESS: *". . . And Have Not Wisdom"*
Daniel Starch, *President*
Daniel Starch & Staff

GENERAL MEETING

"New Developments in the Education of Abler Students"

ACCELERATION: BASIC PRINCIPLES AND RECENT RESEARCH 107
Sidney L. Pressey, *Ohio State University*

COLLEGE ADMISSION WITH ADVANCED STANDING 113
William H. Cornog, *Philadelphia Central High School*

SPECIAL TREATMENT FOR ABLER STUDENTS AND ITS RELATION TO NATIONAL MANPOWER 120
Dael L. Wolfle, *American Association for the Advancement of Science*

APPENDIX 127

GENERAL MEETING

Application of Information Theory to Testing

9

Multiple Assignments of Persons to Jobs*

PAUL S. DWYER

1. Introduction. I want to talk to you about a problem which arises when men are to be assigned to jobs in the most efficient manner. In the time at my disposal I can give you only a brief outline of the nature of the problem, of the general methods proposed for its solution, and of my recent work on the solution of the problem with the use of transformations. However, I am giving you supplementary material which you may examine later in more detail if you desire to obtain a more comprehensive view of the problem and the methods of solution.

2. The Nature of the Problem. A simple illustration may serve to give you some idea of the nature of the problem. A corporation hires 4 college graduates to fill 4 vacancies without determining which individual is to be placed on which job. These graduates, though they have differing abilities as indicated by their records, are all hired for the same salary. The problem of the corporation is then to place these 4 men on the 4 jobs in such a way that the corporation will obtain maximum value from their services.

The corporation may do this by estimating the worth to the corporation in thousands of dollars per year of each individual if he were to be placed in each one of the 4 jobs. Such a set of estimates is shown in Table I. The entries in the table show the values, denoted by c_{ij}, which indicate the estimated contribution to the total effort (in units of \$1000) which individual i will make if he is placed on job j. Thus, individual 1 is most valuable on job 1 but so is individual 2 and individual 4. The problem is to place all 4 individuals on all 4 jobs in such a way that the sum of the assigned c_{ij} values is as large as possible.

Now since each individual can fill one and but one job and since each job must be filled by some one individual, it follows that any assignment of the 4 individuals to the 4 jobs must involve one and but one selection from each row and from each column. Hence the problem becomes one of making selections of c_{ij} values, one from each row and one from each column, so that the sum is as large as possible.

---

* This research was supported in part by the United States Air Force under Contract No. AF18(600)—1050 monitored by Director, Detachment 4 (Crew Research Laboratory), Air Force Personnel and Training Research Center, Randolph Air Force Base, Randolph Field, Texas. Permission is granted for reproduction, translation, publication, use and disposal in whole and in part by or for the United States Government.

A more general statement of the problem results from the display of Table II where the c_{ij} values of N jobs and N individuals are indicated. The problem is to select N values of c_{ij} so that the sum of the c_{ij} values so selected is as large as possible. See references A.

3. Alternate Forms of the Problems. The form of Table I and Table II is that of a square array of c_{ij} values featuring N rows and N columns. As a result of grouping the number of columns may be reduced, in some problems, so that the array takes on rectangular form with height greater than width. For example the problem of Table I may be so reduced since the values of c_{ij} in column 3 are identical with the c_{ij} values of column 4. In so far as the solution is concerned, there is no difference between job 3 and job 4 so that the two jobs may be grouped together in a common job category. If we denote the number of such categories by m, and the number of jobs in job category j by q_j, we have $m = 3$, $q_1 = q_2 = 1$, $q_3 = 2$ for the problem of Table I. The values q_j, which indicate the numbers of individuals to be assigned to the respective job categories, are called quotas. This form of the problem which features these job categories and quotas is sometimes known as the quota form.

The quota form of the problem of Table I is shown in Table III. The quota form of the general problem of Table II is shown in Table IV.

Rows may also be grouped to form personnel categories when the c_{ij} values in different rows are identical or approximately so. When personnel categories and job categories are both used, we have a two-way grouped distribution which takes on the form sometimes called the frequency form. The number of personnel categories is taken as n, and the frequency in personnel category i is indicated by f_i. Of course the sum of the f_i values equals the sum of the q_j values which is N. This frequency form of the problem is illustrated in Table V.

The form of Table I and Table II, since it features the nongrouped values for both individuals and jobs, is sometimes called the individual form.

4. Equivalent Problems. This problem is essentially the equivalent of problems in other fields. For example the Hitchcock transportation problem is the mathematical equivalent of the personnel classification problem though it calls for the selection of the c_{ij} values from each row and column so as to minimize, rather than maximize, the sum. There

is no time here for a discussion of these equivalent problems but references B are provided for those who are interested.

5. Methods of Solution Previously Used. Many methods have been proposed for the solution of this problem. Monographs giving surveys of these different methods are mentioned in references C. Of these methods I should call your attention to the method of all possible assignments, the simplex method, and the method of optimal regions.

In the method of all possible assignments, every possible alternative assignment is made. Now there are N! alternative assignments though some of the assignment sums may be equal. Thus with $N = 4$ in Table I, the computation of the $4! = 24$ possible sums shows us the maximum sum of $23,000. But the method of all possible assignments is impractical for larger values of N since N! increases very rapidly.

The simplex method was designed for more general problems in linear programming and game theory. Dantzig and Votaw have applied it to this problem (see D). It is my thesis that the machinery of the simplex method is unnecessarily complex for this problem and that simpler methods, described below, are preferable.

The method of optimal regions is very useful in solving the classification problem especially when the number of job categories is small (see E). But the method I wish to discuss with you today is the method of transformations by which the array of c_{ij} values can be transformed to a new array from which the solution can be obtained by the selection of zero terms.

6. Method of Transformations. A solution of the problem, and there may be more than one, consists in the assignment of each individual i to some job j so that the sum of the c_{ij} values is to be maximized. This maximum sum is not the solution, though it can be calculated once the solution is known. A solution consists in the assignments, i.e., the pairs of values of i and j. Thus in Table I a solution consists in the assignment of man 1 to job 1, man 2 to job 3, man 3 to job 4, and man 4 to job 2. If we indicate the job assigned to individual i by J_i and consider the men in the order 1, 2, 3, 4 we can write this solution compactly by $J_i = 1, 3, 4, 2$. The solution sum is $6 + 6 + 5 + 6 = 23$ units but this is not the solution. The solution is simply the set of elements (i, J_i).

I can now state an important relation which serves as the basis of the method of transformations. Any constant may be subtracted from every element in any row or column without changing the solution. The solution sum is decreased by the amount subtracted but the solution is not changed. Hence we may subtract simultaneously constants

c_i from every row i and constants c_j from every row j without changing the solution. The purpose of the method of transformations is to make use of successive subtractions from rows and columns until an array results from which the solution is immediately obtainable. Specific directions follow.

The first step in the solution of a maximization problem is the subtraction of the largest element in each row from each element in that row. The results of these subtractions $c_{ij}^{(1)}$, are either 0 or negative. The process is illustrated in Table VI where the maximum values for the rows of Table I are shown at the right of the first array and the values of $c_{ij}^{(1)}$ in the second array. Now the $c_{ij}^{(1)}$ array, since it contains only non-positive terms, cannot have a solution with a sum greater than zero. We cannot locate a solution with sum 0 as long as any column has all non-zero terms. So we subtract the largest element in each column as indicated at the bottom of the array. The resulting $c_{ij}^{(2)}$ array has at least one zero in each row and each column. See the third array of Table VI. Sometimes a solution can be obtained from this array by using only 0 elements. This is not possible in Table VI. An additional transformation is in order.

Before indicating the nature of this transformation, we note that the second and third arrays of Table VI feature negative signs. These could be eliminated if, in the first array, the elements were subtracted from the maximum values rather than vice-versa. Then the subtraction of the smallest element in each column of the second array is indicated. This process is illustrated in the first three arrays of Table VII. The positive elements of the second and third of these arrays are identical, aside from sign, with those of Table VI.

We are unable to find a solution using the 0 elements of the third array of Table VII since the 0 terms in columns 2, 3, 4 are all in row 3. However if we subtract -1 from this row, we can then also subtract 1 (the smallest non-zero value in columns 2, 3 and 4) from each element of columns 2, 3 and 4 to form the $c_{ij}^{(3)}$ array. The net sum of the subtracted values is $1 + 1 + 1 - 1 = 2$ which is placed in the lower right corner. In general any such transformation in which the sum of the subtracted constants is positive and which does not result in negative terms, may constitute the next step of the solution.

In this problem the solution can be obtained from the 0 terms of the $c_{ij}^{(3)}$ array. The 0 terms indicating the solution are marked with asterisks in the $c_{ij}^{(3)}$ array. The corresponding terms in the c_{ij} array are also marked with asterisks. The sum of these terms is the solution sum, 23 units. This may be checked by subtracting from 26, the lower right entry in the first array, the sum of the lower right entries in the two arrays below it. In more complex problems additional transformations of this type may be necessary. But we can prove that it is always possible to make transformations of this type so that the solution may be based eventually on 0 elements. These transformations are fairly easily discovered in practice since they are based on patterns of zeros.

The method is applied to a problem in the frequency form, previously used by Votaw to illustrate the simplex method, in Table VIII. The $c_{ij}^{(2)}$ array results from the subtraction from the maximum row values followed by the subtraction of the minimum column values. But, considering frequencies, there are not enough 0 terms in row 2 and row 3 of the third array to satisfy the frequencies. However, the subtraction of 1 from row 2 and row 3, with the subtraction of -1 from column 3 and column 4 (and a net sum of $30 + 35 - 25 - 28 = 12$) leads to the $c_{ij}^{(3)}$ array with very many 0's. One of the many solutions immediately evident from this array is indicated by superscripts.

A final illustration (Table IX) uses a modification of Brogden's quota form problem with 100 men and 4 job categories. The solution follows the steps outlined above. Subtractions are made from the $c_{ij}^{(1)}$ so as to meet the quotas for column 3 and corresponding subtractions are made from the rows so as to keep one 0 term in each row. A corresponding treatment of the first column of the $c_{ij}^{(2)}$ matrix leads to the $c_{ij}^{(3)}$ matrix with enough 0 terms to reveal the solution indicated in the column headed J_i.

The solution sum 708 units can be obtained by adding the c_{iJ_i} values of the first array indicated by the solution. It can be checked by forming $723 - (6 + 9)$.

7. Concluding Remarks. The method of transformations just described gives a solution to the classification problem which is as simple as one can expect. It can be programmed for machines but, except for the most complex problems, hand methods are quite satisfactory.

We now appear to have a good solution for problems with true c_{ij} values. When true c_{ij} values are unavailable, as they commonly are, questions arise as to the estimation of the values, as to the validity and sampling errors of the estimates, as to the resulting effect on the formulation of the problem, etc. The study of this general area, involving possible alternative procedures using the information available, is very important.

A narrower and more immediate practical problem also commonly confronts us. How can we, with our present knowledge and information, calculate any useful estimates of the c_{ij} to which we can apply the available techniques? (See F.) We might use standard scores of variables correlated with success on the job, as in the Brogden illustration, but, as the vocational counselor knows, such single predictors are not commonly immediately observable. And even combinations of observable predictors, such as those obtained by regression, are not enough for this problem, even if valid, since they must be transformed to estimates of the contribution to the common effort. Some sort of a weight must be given to each particular job since a measure of the importance of the job, as well as the proficiency of the individual on the job, is needed for estimating the contribution of the individual to the common effort. How are we to determine these job weights? Aside from the matter of the validity of the predictors we are forced, for the most part, to rely on the estimates of experts or to use hypothetical weights. In conclusion, I would like to pose this question for future research: How can we use available information in obtaining practical objective measures of those job weights whose determination is prerequisite to any useful solution of the personnel classification problem?

Table I: A Simple Problem

	c_{ij} in $1000 units			
i \ j	1	2	3	4
1	6	3	4	4
2	7	4	6	6
3	3	4	5	5
4	8	6	4	4

Table II: The General Problem

	c_{ij}						
i \ j	1	2	3	. . .	j	. . .	N
1	c_{11}	c_{12}	c_{13}	. . .	c_{1j}	. . .	c_{1N}
2	c_{21}	c_{22}	c_{23}	. . .	c_{2j}	. . .	c_{2N}
3	c_{31}	c_{32}	c_{33}	. . .	c_{3j}	. . .	c_{3N}
. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
i	c_{i1}	c_{i2}	c_{i3}	. . .	c_{ij}	. . .	c_{iN}
. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
N	c_{N1}	c_{N2}	c_{N3}	. . .	c_{Nj}	. . .	c_{NN}

Table III: Quota for Problem of Table I

	c_{ij} in $\$1000$ units		
q_j / i	1	1	2
1	6	3	4
2	7	4	6
3	3	4	5
4	8	6	4

Table IV: Quota Form for General Problem

	c_{ij}						
q_j / i	q_1	q_2	q_3	...	q_j	...	q_m
1	c_{11}	c_{12}	c_{13}	...	c_{1j}	...	c_{1m}
2	c_{21}	c_{22}	c_{23}	...	c_{2j}	...	c_{2m}
3	c_{31}	c_{32}	c_{33}	...	c_{3j}	...	c_{3m}
...	...	...	...	...	...	...	...
i	c_{i1}	c_{i2}	c_{i3}	...	c_{ij}	...	c_{im}
...	...	...	...	...	...	...	...
N	c_{N1}	c_{N2}	c_{N3}	...	c_{Nj}	...	c_{Nm}

Table V: Frequency Form for General Problem

	c_{ij}						
q_j / f_i	q_1	q_2	q_3	...	q_j	...	q_m
f_1	c_{11}	c_{12}	c_{13}	...	c_{1j}	...	c_{1m}
f_2	c_{21}	c_{22}	c_{23}	...	c_{2j}	...	c_{2m}
f_3	c_{31}	c_{32}	c_{33}	...	c_{3j}	...	c_{3m}
...	...	...	...	...	...	...	...
f_i	c_{i1}	c_{i2}	c_{i3}	...	c_{ij}	...	c_{im}
...	...	...	...	...	...	...	...
f_n	c_{n1}	c_{n2}	c_{n3}	...	c_{nj}	...	c_{nm}

Table VI: Successive Transformations for Problem of Table I

	c_{ij}				
j i	1	2	3	4	Max.
1	6	3	4	4	6
2	7	4	6	6	7
3	3	4	5	5	5
4	8	6	4	4	8

	$c_{ij}^{(1)}$			
j i	1	2	3	4
1	0	−3	−2	−2
2	0	−3	−1	−1
3	−2	−1	0	0
4	0	−2	−4	−4
Max.	0	−1	0	0

	$c_{ij}^{(2)}$			
j i	1	2	3	4
1	0	−2	−2	−2
2	0	−2	−1	−1
3	−2	0	0	0
4	0	−1	−4	−4

Table VII: Solution of the Problem of Table I

i	c_{ij}				Max.
1	6*	3	4	4	6
2	7	4	6*	6	7
3	3	4	5	5*	5
4	8	6*	4*	4	8
					26

i	$c_{ij}^{(1)}$				
1	0	3	2	2	
2	0	3	1	1	
3	2	1	0	0	
4	0	2	4	4	
	Min.	1			1

i	$c_{ij}^{(2)}$				
1	0	2	2	2	
2	0	2	1	1	
3	2	0	0	0	−1
4	0	1	4	4	
		1	1	1	2

i	$c_{ij}^{(3)}$				J_i
1	0*	1	1	1	1
2	0	1	0*	0	3
3	3	0	0	0*	4
4	0	0*	3	3	2
					23

Table VIII: Solution of Votaw Problem

	c_{ij}					
f_i \ q_j	15	20	25	28	12	Max.
12	7	8	6	7	8	8
30	6	5	8	6	7	8
35	5	4	5	7	6	7
23	8	9	8	7	9	9
						788

	$c_{ij}^{(1)}$					
f_i \ q_j	15	20	25	28	12	
12	1	0	2	1	0	
30	2	3	0	2	1	
35	2	3	2	0	1	
23	1	0	1	2	0	
Min.	1					15

	$c_{ij}^{(2)}$					
f_i \ q_j	15	20	25	28	12	
12	0	0	2	1	0	
30	1	3	0	2	1	1
35	1	3	2	0	1	1
23	0	0	1	2	0	
			−1	−1		12

Table VIII (*Continued*)

	$c_{ij}^{(3)}$					
f_i \ q_j	15	20	25	28	12	
12	0	0^{12}	3	2	0	
30	0	2	0^{25}	2	0^{5}	
35	0	2	2	0^{28}	0^{7}	
23	0^{15}	0^{8}	2	3	0	
						761

Table IX: Solution of Brogden Problem

	c_{ij}						c_{ij}				
i \ q_j	25	25	25	25	Max.	i \ q_j	25	25	25	25	Max.
1	−8	−24	1	−4	1	51	6	0	−1	−16	6
2	−3	−19	3	7	7	52	0	0	2	−5	2
3	−4	−18	1	3	3	53	3	1	4	−4	4
4	−11	−17	6	11	11	54	5	1	1	8	8
5	−1	−16	−8	−10	−1	55	7	1	0	6	7
6	−5	−15	7	−3	7	56	3	1	4	7	7
7	−7	−14	−1	−13	−1	57	−3	2	−3	−11	2
8	−6	−13	2	4	4	58	−6	2	−7	−3	2
9	−1	−13	−6	−4	−1	59	−1	2	−8	7	7
10	−2	−12	−9	−11	−2	60	4	2	−5	−12	4
11	−7	−12	3	1	3	61	−5	3	4	10	10
12	−9	−11	−7	−14	−7	62	−1	3	3	0	3
13	−3	−11	1	−9	1	63	−3	3	−4	−18	3
14	−8	−10	−3	5	5	64	1	3	2	−9	3
15	−4	−10	6	0	6	65	0	4	−4	−2	4

Table IX continued pages 22–28

Table IX (*Continued*)

16	−3	−10	−5	3	3	66	6	4	1	4	6
17	1	−9	−10	24	24	67	10	4	0	18	18
18	−5	−9	8	9	9	68	4	4	0	0	4
19	2	−9	5	−5	5	69	6	5	−7	−8	6
20	−2	−8	2	12	12	70	4	5	11	13	13
21	−7	−8	6	9	9	71	−2	5	1	−8	5
22	6	−8	−2	9	9	72	9	5	0	−6	9
23	3	−7	2	−1	3	73	−1	6	0	12	12
24	7	−7	−6	2	7	74	0	6	9	19	19
25	−2	−7	3	6	6	75	1	6	4	4	6
26	2	−6	2	8	8	76	8	7	6	−24	8
27	−9	−6	1	−3	1	77	1	7	−4	−1	7
28	−1	−6	−2	−5	−1	78	0	7	1	5	7
29	−6	−5	−11	−17	−5	79	1	8	3	10	10
30	5	−5	−5	−9	5	80	0	8	−5	−3	8
31	2	−5	−1	2	2	81	1	8	−1	0	8
32	−2	−5	0	10	10	82	8	9	0	−1	9
33	−3	−4	3	−7	3	83	−2	9	−2	2	9
34	−6	−4	4	15	15	84	3	9	−6	5	9
35	1	−4	5	3	5	85	0	10	−6	−13	10
36	4	−4	7	1	7	86	2	10	−4	5	10
37	−5	−3	−1	−6	−1	87	5	10	−1	−2	10
38	−4	−3	0	4	4	88	4	11	−2	−12	11
39	7	−3	−1	6	7	89	3	11	−3	−4	11
40	1	−3	−3	−10	1	90	0	12	8	13	13
41	−2	−2	4	17	17	91	9	12	0	11	12
42	−3	−2	3	8	8	92	0	13	−2	−15	13
43	−10	−2	2	−2	2	93	11	13	−1	−1	13
44	−1	−2	7	−7	7	94	1	14	−3	−19	14
45	−1	−1	−3	−2	−1	95	−4	15	−3	3	15
46	2	−1	−9	−6	2	96	2	16	−2	−8	16
47	−1	−1	5	−10	5	97	0	17	−4	2	17
48	−4	−1	10	−5	10	98	3	18	1	14	18
49	2	0	−2	−7	2	99	0	19	1	1	19
50	3	0	5	16	16	100	5	24	0	1	24
					21						723

Table IX (*Continued*)

	$c_{ij}^{(1)}$						$c_{ij}^{(1)}$				
$i \backslash q_j$	25	25	25	25		$i \backslash q_j$	25	25	25	25	
1	9	25	0	5	−2	51	0	6	7	22	
2	10	26	4	0		52	2	2	0	7	−2
3	7	21	2	0		53	1	3	0	8	−2
4	22	28	5	0		54	3	7	7	0	
5	0	15	7	9		55	0	6	7	1	
6	12	22	0	10	−2	56	4	6	3	0	
7	6	13	0	12	−2	57	5	0	5	13	
8	10	17	2	0		58	8	0	9	5	
9	0	12	5	3		59	8	5	15	0	
10	0	10	7	9		60	0	2	9	16	
11	10	15	0	2	−2	61	15	7	6	0	
12	2	4	0	7	−2	62	4	0	0	3	−2
13	4	12	0	10	−2	63	6	0	7	21	
14	13	15	8	0		64	2	0	1	12	−1
15	10	16	0	6	−2	65	4	0	8	6	
16	6	13	8	0		66	0	2	5	2	
17	23	33	34	0		67	8	14	18	0	
18	14	18	1	0	−1	68	0	0	4	4	
19	3	14	0	10	−2	69	0	1	13	14	
20	14	20	10	0		70	9	8	2	0	
21	16	17	3	0		71	7	0	4	13	
22	3	17	11	0		72	0	4	9	15	
23	0	10	1	4	−1	73	13	6	12	0	
24	0	14	13	5		74	19	13	10	0	
25	8	13	3	0		75	5	0	2	2	

Table IX (*Continued*)

26	6	14	6	0		76	0	1	2	32	
27	10	7	0	4	−2	77	6	0	11	8	
28	0	5	1	4	−1	78	7	0	6	2	
29	1	0	6	12		79	9	2	7	0	
30	0	10	10	14		80	8	0	13	11	
31	0	7	3	0		81	7	0	9	8	
32	12	15	10	0		82	1	0	9	10	
33	6	7	0	10	−2	83	11	0	11	7	
34	21	19	11	0		84	6	0	15	4	
35	4	9	0	2	−2	85	10	0	16	23	
36	3	11	0	6	−2	86	8	0	14	5	
37	4	2	0	5	−2	87	5	0	11	12	
38	8	7	4	0		88	7	0	13	23	
39	0	10	8	1		89	8	0	14	15	
40	0	4	4	11		90	13	1	5	0	
41	19	19	13	0		91	3	0	12	1	
42	11	10	5	0		92	13	0	15	28	
43	12	4	0	4	−2	93	2	0	14	14	
44	8	9	0	14	−2	94	13	0	17	33	
45	0	0	2	1		95	19	0	18	12	
46	0	3	11	8		96	14	0	18	24	
47	6	6	0	15	−2	97	17	0	21	15	
48	14	11	0	15	−2	98	15	0	17	4	
49	0	2	4	9		99	19	0	18	18	
50	13	16	11	0		100	19	0	24	23	
			2						2		6

Table IX (*Continued*)

	$c_{ij}^{(2)}$						$c_{ij}^{(2)}$				
q_j / i	25	25	25	25		q_j / i	25	25	25	25	
1	11	27	0	7		51	0	6	5	22	−3
2	10	26	2	0		52	4	4	0	9	
3	7	21	0	0		53	3	5	0	10	
4	22	28	3	0		54	3	7	5	0	
5	0	15	5	9	−3	55	0	6	5	1	−3
6	14	24	0	12		56	4	6	1	0	
7	8	15	0	14		57	5	0	3	13	
8	10	17	0	0		58	8	0	7	5	
9	0	12	3	3	−3	59	8	5	13	0	
10	0	10	5	9	−3	60	0	2	7	16	−3
11	12	17	0	4		61	15	7	4	0	
12	4	6	0	9		62	6	2	0	5	
13	6	14	0	12		63	6	0	5	21	
14	13	15	6	0		64	3	1	0	13	
15	12	18	0	8		65	4	0	6	6	
16	6	13	6	0		66	0	2	3	2	−3
17	23	33	32	0		67	8	14	16	0	
18	15	15	0	1		68	0	0	2	4	−3
19	5	16	0	12		69	0	1	11	14	−3
20	14	20	8	0		70	9	8	0	0	
21	16	17	1	0		71	7	0	2	13	
22	3	17	9	0		72	0	4	7	15	−3
23	1	11	0	5	−2	73	13	6	10	0	
24	0	14	11	5	−3	74	19	13	8	0	
25	8	13	1	0		75	5	0	0	2	

Table IX (*Continued*)

26	6	14	4	0		76	0	1	0	32	−3
27	12	9	0	6		77	6	0	9	8	
28	1	6	0	5	−2	78	7	0	4	2	
29	1	0	4	12	−2	79	9	2	5	0	
30	0	10	8	14	−3	80	8	0	11	11	
31	0	7	1	0	−3	81	7	0	7	8	
32	12	15	8	0		82	1	0	7	10	−2
33	8	9	0	12		83	11	0	9	7	
34	21	19	9	0		84	6	0	13	4	
35	6	11	0	4		85	10	0	14	23	
36	5	13	0	8		86	8	0	12	5	
37	6	4	0	7		87	5	0	9	12	
38	8	7	2	0		88	7	0	11	23	
39	0	10	6	1	−3	89	8	0	12	15	
40	0	4	2	11	−3	90	13	1	3	0	
41	19	19	11	0		91	3	0	10	1	
42	11	10	3	0		92	13	0	13	28	
43	14	6	0	6		93	2	0	12	14	−1
44	10	11	0	16		94	13	0	15	33	
45	0	0	0	1	−3	95	19	0	16	12	
46	0	3	9	8	−3	96	14	0	16	24	
47	8	8	0	17		97	17	0	19	15	
48	16	13	0	17		98	15	0	15	4	
49	0	2	2	9	−3	99	19	0	16	18	
50	13	16	9	0		100	19	0	22	23	
	3						3				9

Table IX (*Continued*)

	$c_{ij}^{(3)}$							$c_{ij}^{(3)}$					
q_j / i	25	25	25	25	j	J_i	q_j / i	25	25	25	25	j	J_i
1	8	27	0	7	3	3	51	0	9	8	25	1	1
2	7	26	2	0	4	4	52	1	4	0	9	3	3
3	4	21	0	0	3,4	3	53	0	5	0	10	1,3	3
4	19	28	3	0	4	4	54	0	7	5	0	1,4	4
5	0	18	8	12	1	1	55	0	9	8	4	1	1
6	11	24	0	12	3	3	56	1	6	1	0	4	4
7	5	15	0	14	3	3	57	2	0	3	13	2	2
8	7	17	0	0	3,4	3	58	5	0	7	5	2	2
9	0	15	6	6	1	1	59	5	5	13	0	4	4
10	0	13	8	12	1	1	60	0	5	10	19	1	1
11	9	17	0	4	3	3	61	12	7	4	0	4	4
12	1	6	0	9	3	3	62	3	2	0	5	3	3
13	3	14	0	12	3	3	63	3	0	5	21	2	2
14	10	15	6	0	4	4	64	0	1	0	13	1,3	3
15	9	18	0	8	3	3	65	1	0	6	6	2	2
16	3	13	6	0	4	4	66	0	5	6	5	1	1
17	20	33	32	0	4	4	67	5	14	16	0	4	4
18	12	19	0	1	3	3	68	0	3	5	7	1	1
19	2	16	0	12	3	3	69	0	4	14	17	1	1
20	11	20	8	0	4	4	70	6	8	0	0	3,4	3
21	13	17	1	0	4	4	71	4	0	2	13	2	2
22	0	17	9	0	1,4	4	72	0	7	10	18	1	1
23	0	13	2	7	1	1	73	10	6	10	0	4	4
24	0	17	14	8	1	1	74	16	13	8	0	4	4
25	5	13	1	0	4	4	75	2	0	0	2	2,3	2
26	3	14	4	0	4	4	76	0	4	3	35	1	1
27	9	9	0	6	3	3	77	3	0	9	8	2	2
28	0	8	2	7	1	1	78	4	0	4	2	2	2
29	0	2	6	14	1	1	79	6	2	5	0	4	4
30	0	13	11	17	1	1	80	5	0	11	11	2	2

TABLE IX (*Continued*)

31	0 10 4 3	1	1	81	4 0 7 8	2	2
32	9 15 8 0	4	4	82	0 2 9 12	1	1
33	5 9 0 12	3	3	83	8 0 9 7	2	2
34	18 19 9 0	4	4	84	3 0 13 4	2	2
35	3 11 0 4	3	3	85	7 0 14 23	2	2
36	2 13 0 8	3	3	86	5 0 12 5	2	2
37	3 4 0 7	3	3	87	2 0 9 12	2	2
38	5 7 2 0	4	4	88	4 0 11 23	2	2
39	0 13 9 4	1	1	89	5 0 12 15	2	2
40	0 7 5 14	1	1	90	10 1 3 0	4	4
41	16 19 11 0	4	4	91	0 0 10 1	1,2	1
42	8 10 3 0	4	4	92	10 0 13 28	2	2
43	11 6 0 6	3	3	93	0 1 13 15	1	1
44	7 11 0 16	3	3	94	10 0 15 33	2	2
45	0 3 3 4	1	1	95	16 0 16 12	2	2
46	0 6 12 11	1	1	96	11 0 16 24	2	2
47	5 8 0 17	3	3	97	14 0 19 15	2	2
48	13 13 0 17	3	3	98	12 0 15 4	2	2
49	0 5 5 12	1	1	99	16 0 16 18	2	2
50	10 16 9 0	4	4	100	16 0 22 23	2	2
							708

REFERENCES

A. To the Nature of the Problem.

1. Brogden, Hubert E. An approach to the problem of differential prediction. *Psychometrika*, 1946, 11, 139-154.
2. Thorndike, Robert L. The problem of classification of personnel. *Psychometrika*, 1950, 15, 215-235.
3. Votaw, D. F., Jr. Methods of solving some personnel-classification problems. *Psychometrika*, 1952, 17, 255-286.
4. Dwyer, P. S. Basic theory and methods in optimal classification of personnel. Personnel Research Branch, Department of the Army, 1953.

B. To Equivalent Problems.

5. Dantzig, George B. Application of the simplex method to a transportation problem. Chapter XXIII. *Activity analysis of production and allocation.* Cowles Commission Monograph No. 13. New York: Wiley, 1951.

6. Rao, C. R. *Advanced statistical methods in biometric research.* New York: Wiley, 1952.

7. von Neumann, John. A certain zero-sum two-person game equivalent to the optimal assignment problem. In H. W. Kuhn and A. W. Tucker (Eds.), *Contributions to the theory of games,* Vol. II. *Annals of Mathematics Study No.* 28. Princeton: Princeton Univer. Press, 1953.

8. Roby, Thornton B. Prerequisites for pair-scores to be used for assembling small work groups. Research Bulletin AFPTRC-TR-54-13 Air Force Personnel and Training Research Center, Lackland Air Force Base, December 1953.

C. To Surveys of Methods of Solution.

9. Smith, Robert B. Hand computational methods for the classification problem. Air Training Command, Human Resources Research Center, September 1951.

10. Votaw, D. F., Jr. and Dailey, J. T. Assignment of personnel to jobs. Research Bulletin 52-24, Air Training Command, Human Resources Research Center, Lackland Air Force Base, August, 1952.

11. Dwyer, P. S. Selected computational methods and related theory in the optimal classification of personnel. Personnel Research Branch, Department of the Army, 1954.

See also (4).

D. To the Simplex Method.

See (3), (4), (5), (10).

E. To the Method of Optimal Regions and Related Theory.

12. Dwyer, P. S. Solution of the personnel classification problem with the method of optimal regions. *Psychometrika,* 1954, 19, 11-26.

13. Lord, Frederick M. Notes on a problem of multiple classification. *Psychometrika,* 1952, 17, 297-304.

See also (1), (2), (6), (11).

F. To the Determination of c_{ij} Values.

14. Dwyer, P. S. Selection and linear combination of tests in relation to multiple criteria and differential classification. PRB Research Note 7, Department of the Army, January 1953.

See also (1), (2).

New Light on Test Strategy from Decision Theory*

LEE J. CRONBACH

In every practical use of tests, our aim is to make decisions. This is obvious in personnel selection and in Dr. Dwyer's assignment problem, but it is also true of testing in the classroom and in the clinic. The teacher uses tests because he has to make decisions about appropriate instructional methods. The clinician uses tests as an aid in deciding on therapeutic tactics. Sometimes, as in vocational guidance, the decisions are made not by the tester but by the person tested. Test theory should indicate how to reach the best possible decision in any of these situations.

We use the word *strategy* to refer to the process by which an individual arrives at a decision. A strategy may be very simple: "I shall examine the applicant's grade average, and if it is B or better I shall accept him." The strategy may instead be complex, stating what tests if any will be given, what decision will be made for any particular pattern of results, and what further steps will be taken to decide on borderline cases. Choosing among alternative strategies is the essential problem of test theory.

There are two questions in choosing a strategy. First, with any given procedure for gathering information, what is the best procedure for translating this information into final decisions? Dr. Dwyer has just shown us the solution for one problem of that type. The second, but logically prior, question is: Among several alternative procedures for gathering information, which is most profitable?

In order to compare two strategies, we have to determine how much benefit we gain from either one. Most of the problems of decision theory therefore reduce to determining just how much benefit is gained from a particular decision-making procedure.

Since this morning's program is intended to deal primarily with insights from some of these newer points of view, I shall not dwell on the mathematics of decision problems. There is available a large amount of relevant theory in the work of economists on utility, in the theory of games, and in the statistical decision theory of Abraham

---

* Based on work conducted under Contract N6ori-07146 with the Office of Naval Research.

Wald. Decision problems can be attacked in many ways, but we have confined ourselves to strategies which maximize expected utility. This is reasonable only if we are dealing with a stable and familiar situation.

The decision model requires us to specify three aspects of any decision. One is the proposed strategy or decision rule. For example, the strategy might be to give two tests, combine scores by a regression formula, and accept everyone above a given cutoff. Second, we consider the adequacy of the information to be used. The usual contingency matrix or scatter diagram relating test scores and criterion scores deals with this question. Dr. McGill's work with information analysis is primarily concerned with studying validity matrices. The third necessary element is an evaluation matrix. This, sometimes called a payoff matrix, states specifically just what benefit or detriment accompanies each possible decision. Dr. Dwyer's Tables I and II are evaluation matrices (but his problem is so stated that the validity of the predictors also affects his entries). Once these three aspects of a problem have been described, we are ready to compute the payoff a person can expect if he bases decisions upon this information and this strategy.

Dr. Dwyer properly drew attention to the fact that it will be difficult and at times impossible to write the evaluation matrix for a particular situation. To let this difficulty deter us from using decision theory, however, would be to deny the possibility of sound test theory. Test effectiveness simply *cannot* be evaluated without an evaluation matrix. Even the conventional procedures of test analysis assume certain payoffs covertly, and the reasonableness of some of these hidden assumptions is open to question. In the future, testers may wish to determine utilities by a sort of cost accounting in any specific practical situation, in order to arrive at the best decisions. Our project is proceeding along different lines. We are working with hypothetical (but we hope realistic) decision problems. By assuming that the evaluation matrix has some characteristic form, we are able to judge the utility of different types of tests and strategies. Such an approach can be no better than our assumptions. We hope nonetheless to arrive at general principles of testing which will illuminate many real situations.

Let me turn to some of the concepts a decision approach brings to our attention. I shall cover four such points.

1. Our model suggests that the value of test information should be judged by how much it improves decisions over the best possible decisions made without the test, whereas the conventional validity coefficient reports how much better test decisions are than *chance* de-

cisions. In the majority of situations where tests could be used, a substantial amount of information is already available, and if no test were given the decision would still be considerably better than chance.

Our most valid tests are essentially work samples of their criteria. Where such a work sample might be used, evidence of past performance is also a valid basis for decisions, and such evidence is often readily available. In predicting school marks, for example, a scholastic aptitude test is not greatly more valid than past school records. The contribution of this test to decision making is much smaller than its zero-order validity coefficient would indicate, because better-than-chance decisions could be made without it.

A similar conception applies to classroom testing. The basic knowledge and skill objectives can be assessed with considerable accuracy from day-to-day assignments; a test can add only a small increment to the soundness of decisions. On the other hand, a teacher has rather little basis for judging which pupils have problems of adjustment. The teacher may therefore gain more useful knowledge from a test of adjustment which has limited validity, than from an achievement test which largely duplicates data already available. There are serious weaknesses in our tests for such educational outcomes as creativity, reasoning habits, attitudes, and application of knowledge to problem situations. They are markedly inferior in validity to tests of general intelligence or factual knowledge. But the factors that make testing difficult also prevent valid non-test decisions about these objectives. It may therefore be wiser to use imperfect tests of important objectives that are hard to measure, than to use highly valid tests that merely supplement non-test data.

2. Utility analysis leads us to examine the value of adapting to individual differences in either selection or placement. This can best be considered in terms of a placement problem, such as assigning students to sections of freshman English according to their initial ability. We might think of the various levels as predetermined, and of the test as assigning persons to each category using fixed cutting scores. It is sounder to see the test, the curricula, and the cutting scores as interlocked. We can increase or decrease the demands instruction in any section makes, to fit it to the ability of the persons assigned. Under this procedure we benefit more from testing than when we leave the treatment fixed.

Certain simple assumptions lead to interesting conclusions. If a sample is divided into groups, using fixed cutting scores, the extent to which treatment for the groups should be differentiated depends on the validity of the placement test. If the information has zero validity, utility is maximized when we teach all sections in the manner suited to the average of the population. As validity increases, the treatments given the sections may differ more; but no matter how valid the test, there is an optimum degree of differentiation of treatment. If treatment is differentiated beyond this point, the benefit from sectioning declines. Indeed, it is possible to differentiate treatments so radically that a *loss* in utility results from sectioning even though the test used has considerable validity.

This analysis raises serious question as to whether we are right when we urge teachers to adapt to individual differences. If the teacher has a standard plan, well fitted to the average of the group, he should hesitate to depart from it. Marked alteration of the plan to fit individuals appears to be advisable only when individual differences are validly assessed and their implications for treatment clear.

3. We turn, now, to another suggestion encountered in decision theory. It is customary to look at a test as a unit, and to use it just for one terminal decision. At any point in testing, however, we can make a terminal decision or can continue to gather information. New frontiers open for us when we view testing as a multi-stage, or sequential, operation.

Suppose, in a simple selection problem, we have several short aptitude tests, which together might constitute a selection battery. We give the first short test; some men can be rejected or accepted at once, but less clear-cut cases are retained for further testing. After the second test is given to these men, we can make more final decisions, and only a borderline group goes on to the third test. This process terminates when the benefit from information to be gained at any stage is outweighed by the cost of testing. Considering cost of testing, the sequential method is more profitable than giving the same test to everyone. If testing is expensive, one reaches the final decision for a surprisingly large proportion of men, after only the first short test. One paper on this line of attack has been published by Arbous and Sichel (1), but our detailed results will differ from theirs in important respects.

A sequential plan would require new ways of organizing testing. I shall discuss one procedure for possible use in vocational guidance, clinical diagnosis, or evaluation of classroom learning. Here it can be described in terms of the job assignment problem. For different jobs, so many abilities are relevant that we cannot hope to measure them all accurately in a reasonable period. With brief tests, however, one could crudely measure as many as 50 variables in a half day. Such a survey will indicate some jobs for which the man is an unlikely prospect, and a second group of jobs for which the tests show possible high aptitude. For the man's second test session, perhaps only an hour after the first test, we would assemble a set of booklets to test him more thoroughly on those promising aptitudes. This progressive narrowing would be continued. When the final job assignment is made, we would have a highly reliable measure of the man's aptitude for that job, and also a good measure for the other jobs seriously considered as alternatives. But we would have wasted little time in getting an accurate measure of his space ability or his dexterity, if these areas were not among his better aptitudes on the first survey.

We might actually develop different sorts of tests for the earlier and later screens. The Strong Vocational Interest Blank might be replaced by a brief questionnaire, perhaps one page long. This seems likely to identify the important interest groups for a given man. There might then be a separate, longer interest blank for each of these interest groups, to provide more precise differentiation between related occupations than is now possible.

4. Perhaps our most far-reaching conclusion is that we should take a more favorable view of tests with low validity. Traditionally, if a score has low validity, we conclude that it should not be used. But such tests become valuable when selection ratios are low (as Taylor and Russell noted), when they give even a little new evidence on an important decision, and particularly when they are used as a preliminary survey.

The survey is especially important when many decisions are to be made. Sometimes, as in vocational choice, the decisions are interrelated and lead to one final course of action. The decisions may, on the other hand, be quite independent, as when one diagnoses many persons. The problem in testing is ordinarily to select information-getting devices which will yield greatest benefit for the time available. If we spend a lot of time to get an accurate answer to one question, we must answer other questions without added information. In this situation it may be much wiser to use several tests of limited validity, so that every decision is made with some wisdom, than to get highly accurate information for just one of the decisions.

The difference between validity and utility is clear when we compare group and individual tests. An individual mental test measures one person, with essentially the same expenditure of effort by the tester as the group test measuring one hundred. If the two tests have the same validity, the group test gives us 100 times as much information, and bears on 100 decisions while the individual test bears on one. Hence the improvement of decisions is vastly greater when the group test is used. If the individual test has much higher validity than the group test, which is best to use would depend on the specific decision problem.

This conception permits a favorable view of interviews and other clinical procedures which cover many aspects of the personality. These methods, though undependable, are well suited to a wide-casting survey, gathering a little information on each of hundreds of questions. Such a preliminary scanning draws attention to the critical areas where further information should be gathered prior to any decision. The traditional, narrowly-focussed measuring device is ideal when we know in advance exactly what question needs to be answered. But in deciding whether a man will make a good executive, or in locating a patient's chief conflict areas, no such focus is possible. The first task of the assessor is to discover which critical variables will dictate the proper decision about the individual; in different cases, different variables will be critical.

Personnel workers regard the interview as indispensable, and clinicians have considerable faith in projective methods and qualitative analysis of intelligence test protocols. In my opinion, this faith has developed largely because of rewarding experience with these techniques in their survey function, i.e., as the first stage in a sequential assessment. If it is true that these multi-dimensional techniques have a unique place in assessment, we should judge how well they do that job, and should not demand that they be good measuring instruments —which they are not. On the other side of the picture, if their proper function is to make preliminary surveys so that more intensive examination can follow, one should not rest final judgments on these fallible instruments.

Taken as a whole, decision theory is a mathematical system which permits us to examine the problems we face in developing tests, choosing between tests, and interpreting tests. Whenever we can specify any particular decision problem in the detail Professor Dwyer's problem required, then decision theory can tell us just what to do. By studying common type-problems, decision theory can also offer general recommendations regarding testing strategy.

Conventional test theory assumes that we use tests to obtain numerical measures on an interval scale, as in the physical sciences. That is rarely or never true. The function of psychological and educational tests is to aid in making discrete decisions. The greatest contribution of decision theory is to help testers see this function more clearly.

(See next page for references)

REFERENCES

1. Arbous, A. G., and Sichel, H. S. On the economies of a pre-screening technique for aptitude test batteries. *Psychometrika*, 1952, 17, 331-346.
2. Blackwell, David, and Girshick, M. A. *Theory of games and statistical decisions*. New York: Wiley, 1954.
3. Brogden, H. E., and Taylor, E. K. The dollar criterion—applying the cost accounting concept to criterion construction. *Personnel Psychol.*, 1950, 3, 133-154.
4. Bross, Irwin D. J. *Design for decision*. New York: MacMillan, 1953.
5. Cronbach, Lee J. The counselor's problems from the perspective of communication theory. In Vivian H. Hewer (Ed.), *New perspectives in counseling*, Minnesota Studies in Personnel Work, No. 7. Minneapolis: Univer. of Minnesota Press, 1955.
6. Edwards, Ward. The theory of decision making. *Psychol. Bull.*, 1954, 51, 380-418.
7. Goodman, Leo A. The use and validity of a prediction instrument. *Amer. J. of Sociol.* 1953, 58, 503-512.
8. Thrall, R. M., Coombs, C. H., and Davis, R. L., Editors, *Decision Processes*, New York: Wiley, 1954.

The Relation Between Uncertainty and Variance

WILLIAM J. McGILL

There has been a great deal of hue and cry about information theory in psychological circles. In the midst of this hue and cry, it is easy to become confused about where information theory is supposed to make contact with psychology. What, if anything, should information theory contribute to psychology?

The theory deals with transmitting symbols from one system to another. That certainly sounds familiar. Many questions that come up in learning, in perception, psycho-physics, testing (just to mention a few areas), take on new significance in the terms of information theory. Consequently when we read presentations of information theory and run across words like "message," "noise," "channel capacity" our nostrils begin to quiver and we sniff a familiar scent in the air. At once we try to find analogues for the classical problems of psychology in the theorems presented to us by the communication engineers. This is one large zone of contact between information theory and psychology. Information theory is a source of analogies and ideas that might not have occurred to us, if we thought about our problems in another way. Perhaps the analogies are helpful, perhaps not. I would rather not discuss the merits of using information theory for this purpose. I bring it up only because I want to hammer at the distinction between information *theory* and information *measures*. The theory is concerned with transmitting symbols despite noise. Information *measures* are concerned with the arithmetic of mean-log-probability.

Mean-log-probability is something like variance. It measures the amount of spread in a discrete probability distribution. The formula for mean-log-probability is usually written as follows:

$$U(y) = -\sum_{k=1}^{k=r} p(k) \log_2 p(k)$$

In this formula, y is a variable that can assume any one of *r* discrete values. Each of these values has some probability, p(k). The negative sign before the summation insures that U(y) is positive. The interesting thing about mean-log-probability is that any numbers or measurements which the various categories of y might represent do *not* appear in the formula for U(y). In other words U(y) is non-metric.

Mean-log-probability can be calculated for a variable like "methods of parental discipline" (just to choose an example), where no numbers can be attached to any of the methods. One method is as good or as bad as another. All you can say about them is that they are different. Obviously, you can also compute mean-log-probability for metric variables like IQ, but in so doing you sacrifice whatever you might gain from knowing that an IQ of 100 is very close to an IQ of 105 and very far from an IQ of 150.

In the formula for mean-log-probability, logs are taken to the base 2. This is done in order to provide a simple unit—the "bit." When U(y) is measured in bits, it turns out that U(y) is the average number of binary, or two-category decisions required in order to identify one of the values of y exactly. It is not really necessary to measure mean-log-probability in bits, but almost everyone does.

You can apply mean-log-probability in many situations where information theory has nothing whatever to say, and where in fact an attempt to apply information theory might even look a little ridiculous.

For example, suppose you want to express the relation between anxiety level in children and several methods of parental discipline. The relation might be demonstrated very easily by using mean-log-probability. But our understanding of the relation would not be enhanced if we considered the disciplinary methods as messages and the anxiety levels as the versions of these messages received at the end of the channel. The "capacity" of such a "channel" is of little interest to us. The fact that we can with effort construct information-theory-type interpretations of relations like this one, merely demonstrates that communication engineers and psychologists measure things with roughly the same kind of arithmetic. It does not mean that information *theory* has anything significant to contribute to our understanding of the relation between discipline and anxiety. Consequently we can be interested in the information measures entirely apart from their significance in information theory. To fortify this distinction, let us now call these information measures by the name "uncertainty" measures. Uncertainty means mean-log-probability. It has no necessary connection with information *theory*.

In this paper I want to show that uncertainty is essentially nonmetric variance. This can be a very hollow claim revealing little more than a superficial similarity unless we are prepared to outline in detail the properties of variance that uncertainty possesses.

What are the principal properties of variance?

1. Variance is a measure of spread or variability—so is uncertainty.
2. Variances that are independent are additive—uncertainties that are independent are also additive.

3. Variances can be partitioned into components that reflect contributions to the total variance from a number of predicting variables taken one at a time and in combination; uncertainty has precisely the same property.

4. Variances are seriously affected by the metric or scale properties of the data from which they are derived. On the other hand uncertainty is unaffected by the metric or scale properties of the data. You can chop a distribution into parts and rearrange all the parts without changing uncertainty—but the variance *will* change radically.

It is obvious that uncertainty and variance are closely related but it is equally obvious that the two measures are not identical. There is no simple equation that takes you from uncertainty to variance. Instead, the parallel is in terms of structure or operations. To make this point clearer I have prepared two tables, Tables 1 and 1A in the handout. Table 1 shows the symbols, formulas, and definitions used in a double-classification analysis of variance. Table 1A shows the same arrangement for a double-classification uncertainty-analysis. You can see by glancing across the tables that the parallelism is very complete. In both cases the predictor variables do not have to be metric. They can be pure classifications like our example of methods of parental discipline. In the variance analysis the dependent variable is metric. It is something that can be measured numerically. I have pictured the dependent variable also as having discrete classifications but this is just a convenience of notation.

You can see the relation between the two analyses very clearly when you look at this pair of equations:

$$U(y) = U_{wx}(y) + U(y{:}w,x) \quad (1)$$

$$V(y) = V_{wx}(y) + V(y{:}w,x) \quad (1a)$$

The definitions of the terms in these equations are explained in Tables 1 and 1A. The equations state that the variability of the criterion or dependent variable can be analyzed into predictable and unpredictable parts. Furthermore the predictable variability may be decomposed as follows:

$$U(y{:}w,x) = U(y{:}w) + U(y{:}x) + U(y{:}\overline{wx}) \quad (2)$$

$$V(y{:}w,x) = V(y{:}w) + V(y{:}x) + V(y{:}\overline{wx}) \quad (2a)$$

where again U stands for uncertainty and V stands for variance. These equations state that the total predictable variability can be broken down into a part predictable from w, a part predictable from x, and a part predictable from unique combinations of w and x.

Sometimes the uncertainty interaction term, $U(y{:}\overline{wx})$, can be negative. What is not generally realized is that $V(y{:}\overline{wx})$, a component of *variance*, can also be negative if it is defined as it is in Table 1. The reason is that this term contains a hidden cross product, a correlation term, as well as a sum of squares. Furthermore the correlation term has a negative sign and can sometimes be greater in magnitude than the interaction sum of squares. This means that the interaction term in the variance-analysis can sometimes be pushed negative by the negative correlation term.

When does this happen? It happens when the predictor variables are not independent. In ordinary analysis of variance this situation never comes up because we are careful to construct the analysis so that there are equal numbers of observations in every cell. This automatically insures that the predictor variables are independent. We can guarantee that things behave properly in both uncertainty-analysis and variance-analysis by making the predicting variables orthogonal, i.e. independent. When the predictor variables have some sort of metric this is equivalent to requiring that the correlation between the predictors is zero. In fact, all the arguments I have been making for the similarity between uncertainty-analysis and analysis of variance are equally true of multiple regression analysis, i.e. the case in which the variances in Table 1 are computed around a regression plane. In multiple regression any interaction computed by the rule shown in the last line of Table 1 is bound to be due to non-orthogonality.

We see that multiple correlation, analysis of variance and uncertainty analysis are all very closely related. The different predictive methods are necessitated because sometimes the predicting variables have metrics and sometimes not, sometimes the dependent variable has a metric and sometimes not.

What remains stable in each case is the *structure* of the statistical process of prediction, the operations involved in making the prediction.

One important practical consequence of this invariance of structure in statistical prediction is that we can now analyze non-metric data in the manner of analysis of variance without resorting to the distortion of giving the data an artificial metric. The non-metric analysis is carried out with uncertainty measures and mean-log-probability. Another very important consequence is that all the literature on experimental design in variance analysis is directly applicable to the uncertainty analysis.

Table 1

Symbols, formulas, and definitions used in double classification analysis of variance. The criterion variable, y, is assumed to be metric. The predictor variables, w and x, are categorized but not necessarily metric.*

Symbol	Formula	Definition
1. $V(y)$	$\sum_k \frac{n_{..k}}{n}(y_k - \bar{y})^2$	*Total Variance:* The variance of the criterion variable y.
2. $V_w(y)$	$\sum_i \frac{n_{i..}}{n} \sum_k \frac{n_{i.k}}{n_{i..}}(y_k - \bar{y}_{i..})^2$	*Conditional Variance:* The variance of y when one predictor variable, w, is held constant.
3. $V_{wx}(y)$	$\sum_{i,j} \frac{n_{ij}}{n} \sum_k \frac{n_{ijk}}{n_{ij}}(y_k - \bar{y}_{ij})^2$	*Error Variance:* The variance remaining in y when both predictor variables, w and x, are held constant.
4. $V(y{:}w)$	$V(y) - V_w(y)$	*Main Effect:* The variance of y due to the predictor variable w.
5. $V_w(y{:}x)$	$V_w(y) - V_{wx}(y)$	*Partial Main Effect:* The variance of y due to the predictor variable x, when the predictor variable, w, is held constant.
6. $V(y{:}w,x)$	$V(y) - V_{wx}(y)$	*Total predictable Variance:* The variance of y due to the joint influence of the predictor variables w and x.
7. $V(y{:}\overline{wx})$	$\begin{cases} -V(y) + V_w(y) + V_x(y) - V_{wx}(y) \\ V_w(y{:}x) - V(y{:}x) \end{cases}$	*Interaction variance:* The variance of y due to unique combinations of the predictor variables, w and x.

---

* The criterion variable is indicated by y. It can assume any value y_k. The two predictor variables are w and x, and these two variables can assume values w_i or x_j. Thus the data arrays are three dimensional. In the three dimensional data matrix, n_{ijk} refers to the number of cases occurring in a single cell. The dot notation indicates that n_{ijk} has been added over the missing subscripts.

Table 1A

Symbols, formulas, and definitions used in three-variable uncertainty analysis. The criterion variable y is assumed to be non-metric. The predictor variables, w and x, are categorized and may or may not be metric variables.*

Symbol	Formula	Definition
1. $U(y)$	$-\sum_{k} \frac{n_{..k}}{n} \log_2 \frac{n_{..k}}{n}$	*Total uncertainty:* The amount of uncertainty in the criterion variable y.
2. $U_w(y)$	$-\sum_{i} \frac{n_{i..}}{n} \sum_{k} \frac{n_{i.k}}{n_{i..}} \log_2 \frac{n_{i.k}}{n_{i..}}$	*Conditional uncertainty:* The amount of uncertainty in y when one predictor variable, w, is held constant.
3. $U_{wx}(y)$	$-\sum_{i,j} \frac{n_{ij.}}{n} \sum_{k} \frac{n_{ijk}}{n_{ij.}} \log_2 \frac{n_{ijk}}{n_{ij.}}$	*Error uncertainty:* The amount of uncertainty remaining when both predictor variables, w and x, are held constant.
4. $U(y{:}w)$	$U(y) - U_w(y)$	*Contingent uncertainty:* The uncertainty in y due to the predictor variable, w.
5. $U_w(y{:}x)$	$U_w(y) - U_{wx}(y)$	*Partial Contingent Uncertainty:* The uncertainty in y due to the joint influence of the predictor variables w and x.
6. $U(y{:}w,x)$	$U(y) - U_{wx}(y)$	*Multiple Contingent Uncertainty:* The uncertainty in y due to the joint influence of the predictor variables, w and x.
7. $U(y{:}\overline{wx})$	$\begin{cases} -U(y) + U_w(y) + U_x(y) - U_{wx}(y) \\ U_w(y{:}x) - U(y{:}x) \end{cases}$	*Interaction uncertainty:* The uncertainty of y due to unique combinations of the predictor variables, w and x.

DISCUSSION

PARTICIPANTS

Lee J. Cronbach, Edward E. Cureton, Paul S. Dwyer, Dorothea Ewers, M. M. Kostick, William J. McGill, Victor H. Noll, Joseph Zubin

Dr. Noll: This is by way of comment rather than criticism on a point that Dr. Cronbach made in his paper. I have reference to the second point in relation to use of tests in providing for individual differences.

It seems to me that the implication there was a little unrealistic in that we do not ordinarily, in the classroom, attempt to provide for individual differences on such a mechanical basis. We do not say, "Your IQ is 117; consequently, you do this," and "Your IQ is 90, and you do this."

It seems to me it is more a matter of providing a variety of materials from which the student then, by a process of selection, determines what is suitable for him. In other words, it is more a curricular problem than it is a problem of measurement. Perhaps I am putting an implication on Dr. Cronbach's statement which he did not intend, but I think it is important that we think of it in terms of providing a variety of experiences from which students of different types of ability and interest then choose rather than a mechanical process of determining, on the basis of a test, just what a student is going to do.

Dr. Cronbach: I agree with Dr. Noll's views on the curriculum. The generalization of the paper is this: our system of analysis forces us to reconsider much of the doctrine we have had regarding individual differences. This doctrine takes a variety of forms. For instance, experiments on homogeneous groups have not led to conclusive results. In these studies, pupils have been given different treatments, treatments varied rather mechanically, plus some additional flexibilities such as Dr. Noll mentioned. The results to be expected from X-Y-Z sectioning will differ greatly, depending upon the extent to which X pupils truly differ from Y's and Y's differ from Z's. I know of nothing done in these research programs to make sure that the adaptation to individual differences applied in each section was optimal.

If treatment is overdifferentiated or underdifferentiated, they certainly do not get the full value of homogeneous grouping.

Our result also bears on the assessment literature. Thus it has been said that there must be something wrong with counselors because their standard deviation of estimates of ability is lower than the measured standard deviation. But the evidence is that counselors are making mistakes chiefly because they are *over*differentiating, taking the tests too seriously.

Obviously we have been encouraging teachers to look at the students' personalities and to handle some of them differently from others, on the basis of judgment that one needs encouragement and another needs stern treatment.

The answer lies in this direction: it is fine to differentiate treatment, so long as you do it on a tentative basis, and allow for trial and error. If you make irreversible decisions, such as "you cannot go to college," then your information has to be extremely good.

Dr. Cureton: It seems to me we might consider one illustration while we are on this topic, because it happens to be fairly common in American universities and to rest on pretty definite information of a non-metric type. This is the matter of freshman English. I think probably in colleges and universities you will find more specific sectioning in English than in almost any other area. The reason is that there seem to be some fairly definite cutoffs which imply specifically different types of instruction. For example, if a student can write good English sentences quite regularly, and can also handle them in paragraphs and larger units with acceptable style, many universities will excuse them from freshman English at the outset. Secondly, if he can write acceptable grammatical sentences, but does not do too well in organizing them into paragraphs and larger units and produces poor style, he is likely to be assigned to a regular section. And finally, if he is unable to write grammatical sentences and handle properly the basic elements of grammatical structure, then he is likely to be assigned to a special section in English in which he will get the type of drill that is not needed by those whose secondary preparation is somewhat better.

Dr. Cronbach: Our analysis suggests research on the performance of students of differing initial ability under various treatments along this continuum, from very routine English training to the highest level of English. The highest level might involve going into matters of style, for instance, well above the usual considerations of clarity.

Dr. Kostick: This is for Dr. Cronbach: I was wondering about using the sequential method in entrance examinations.

I carried out a study of the effectiveness of entrance examinations in our Massachusetts State Teachers Colleges. We found that by the end of the sophomore year, of those people who were admitted by certification, only 6% "flunked out" or dropped out with low grades. We also found that for those people who were admitted with an entrance examination mark of over 160, only 18% were dropped. Of course, the mark of 160 or any other mark doesn't have meaning unless you know what it stands for. We used a combination of the score on the psychological examination and the score on the English Examination.

What I am particularly interested in, is the lowest group. We have a range between 130 and 140 which is questionable. From that group, 48% "flunked out."

I was wondering if it would be a good suggestion to see whether we could use some sort of sequential testing of these people, instead of going to the expense of putting them through the freshman or sophomore year and then having practically 50% of them drop out.

Dr. Cronbach: In sequential testing you often can get more thorough measures than you could afford to give to everyone. There is no compelling reason why borderline cases could not sit for a two-day examination even though you wouldn't think of doing this with every entrant. However, you are likely to make a great deal more gain if you find some new sources of tests that will supplement your prediction rather than just extending the old one.

The statement that was made focuses on the importance of having very clearly in mind your value assumptions before you try to make decisions about test strategy. We can each make our own value assumption. I would be hesitant in this particular example to accept the implication that if a person flunks out after two years, this result has negative utility. There may be positive utility in those two years of college, particularly when judged from the viewpoint of the individual, rather than from the values of the institution. I think we fall too easily into seeing the problem only from the institutional side.

Dr. Ewers: I simply want to add to Dr. Cureton's statement that we do the same thing for freshmen in high school, except we have one more category, that is, if they cannot read, we put them in a fourth group. But I wonder really how well we do teach them in these various levels. If we could do a better job there, we might not have so much trouble at the college level.

Dr. Zubin: Dr. McGill, I am a little puzzled by your analysis, especially by the terms you used. They evoke memories of previous knowledge and I wonder whether you would like to connect up for us the previous meaning attached to these times with the meaning you have given them. Analysis of variance gives the components of the total variance in a systematic fashion, and, of course, leaves a little bit for the uncontrolled variance as an error term. Ordinarily we refer to it as the standard error, something left with which to measure the significance of the single components. As a matter of fact, what is the relationship between uncertainty and these older ideas of standard error? Are they at all related, or are they two completely independent things?

Dr. Dwyer's method, of course, applies only to situations where you already know the value of each job and you know the value of the person, and so on. Now, could that method be applied to situations where we still do not know these values? In the clinical field, for example, we do not know the amount of money we could save by a certain procedure. Would it be possible to set up a contingency table with unknown c_{ij}'s and then solve for them under certain conditions? Rather than assume you know your c_{ij}'s, could you put in specified c_{ij}'s and see what happens?

Dr. McGill: The answer is that it was my intention to try to evoke associations from as many people as possible, because I believe that if you have the old associations firmly in hand, it is extremely easy to manipulate these so-called new concepts. They are not new at all.

The analogy of uncertainty with error variance is an example. The "error" uncertainty contains error plus everything else you forgot to analyze. If you have constructed an experiment properly (which almost nobody ever does), the error will be what the model claims it is.

The interesting thing about uncertainty analysis is its technique for testing null hypotheses. In this analysis you do not test the predictable components against error; you test the error against zero and the predictable components against zero. It is a tricky little switch, but the interpretation of the components is identical with variance analysis.

Dr. Dwyer: You can't work the problem unless you have some values of c_{ij}. There isn't any reason why you can't have a whole series of hypothetical c_{ij}'s and maybe the collective problem would be some sort of answer. This formulation of the problem demands it, but we could have a whole series of hypothetical values, which might be interesting.

SECTION 1

Recent Advances in Psychometric Methods

Some Recent Results in Latent Structure Analysis

T. W. ANDERSON

Latent structure analysis may be thought of as an analysis of discrete data that is analogous to factor analysis of continuous data. The usual model for factor analysis is

$$(1) \qquad x_i = \sum_{\nu=1}^{m} \lambda_{i\nu} f_\nu + \mu_i + u_i, \qquad i = 1, \ldots, p,$$

where x_i is the i-th test score, f_ν is the ν-th factor score, μ_i is a constant and u_i is the sum of the i-th specific factor and the error of measurement. The test scores are observed; the quantities on the right hand side are unobserved or latent.

For convenience of discussion, we shall assume that the f_ν and u_i are normally distributed. The key assumptions of this model are that the set u_i are statistically independent of (uncorrelated with) the set f_ν and the u_i are statistically mutually independent (mutually uncorrelated). This means that if we take the subpopulation defined by the requirement that the factor scores are given values, then in this subpopulation the test scores are statistically independent; that is, given the factor scores the predictability of one test score from another is zero. More specifically, given the values of f_ν, the test scores are normally and independently distributed with means $\Sigma_\nu \lambda_{i\nu} f_\nu + \mu_i$ and variance σ_i^2. These assumptions are reflected in the formulas for the variances and covariances of the test scores. If we assume the factors are uncorrelated and have unit variances and let the mean of u_i be zero and the variance of u_i be $E\, u_i^2 = \sigma_i^2$, then the variance of the i-th test score is

$$(2) \qquad E(x_i - \mu_i)^2 = \sum_{\nu=1}^{m} \lambda_{i\nu}^2 + \sigma_i^2,$$

and the covariance of two test scores is

$$(3) \qquad E(x_i - \mu_i)(x_j - \mu_j) = \sum_{\nu=1}^{m} \lambda_{i\nu} \lambda_{j\nu}, \qquad i \neq j.$$

We can say that the common factors "explain" the interdependence or correlation of the test scores; the effect of the u_i appear only in the variances.

In considering latent structure analysis, we shall assume that each item is a dichotomy; that is, an item score takes on the value one for a positive response and zero for a negative response. Thus the continuous test scores x_i are replaced by the discrete item scores y_i. The continuous factor scores f_ν are replaced by the latent attributes g_ν, which may be discrete or continuous. In the subpopulation defined by given latent attribute scores, the responses to the p items are statistically independent; in other words, given the latent attribute score the predictability of one item from another is zero. In such a subpopulation let $\pi_i(g)$ be the probability of a positive response to the i-th item; for ease of exposition we shall assume there is only one latent attribute score g. Then, for example, the probability of a positive response on every item is

$$(4)\quad \Pr\left\{y_1 = 1, y_2 = 1, \ldots, y_p = 1 \mid g\right\} = \pi_1(g)\pi_2(g)\ldots\pi_p(g).$$

For given g, the y_i are a set of independent binomial variables. Next we assume that there is a distribution of the latent attribute, say f (g) over the whole population. The probabilities of $y_1, \ldots, y_p$ or the relative frequencies of various response patterns for the entire population are obtained from those for the subpopulations by averaging with respect to f (g).

To make these ideas more concrete we shall consider two special cases of the latent structure model. One of these can be derived from the factor analysis model. For simplicity we shall assume that there is one common factor f and $\mu_i = 0$; that is,

$$(5)\qquad x_i = \lambda_i f + u_i, \qquad i = 1, \ldots, p.$$

Suppose that the continuous score x_i is replaced by a dichotomous variable y_i with $y_i = 1$ if $x_i > a_i$ and $y_i = 0$ if $x_i \leq a_i$. From (5) we know that $x_1, \ldots, x_p$ are normally distributed with means zero, variances $\lambda_i^2 + \sigma_i^2$ and covariances $\lambda_i \lambda_j$, and from this fact we can compute the probability of a pattern of the dichotomous responses. For example,

$$(6)\ \Pr\{\, y_1 = 1, \ldots, y_p = 1 \,\} = \int_{\alpha_p}^{\infty} \cdots \int_{\alpha_1}^{\infty} h(x_1, \ldots, x_p)\, dx_1 \ldots dx_p,$$

where $h(x_1, \ldots, x_p)$ is the density of the normal distribution. When f is fixed, the conditional distribution of x_i is normal with mean $\lambda_i f$ and variance σ_i^2; in this subpopulation the probability of a positive response on the i-th item is

$$(7)\quad \Pr\{ y_i = 1 \mid f \} = \pi_i(f) = \int_{\alpha_i}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma_i} e^{-\frac{1}{2}(x_i - \lambda_i f)^2/\sigma_i} dx_i = 1 - \Phi\left(\frac{\alpha_i - \lambda_i f}{\sigma_i}\right),$$

where $\Phi(b)$ is the probability of a unit normal variable being less than b. From the fact that the u_i are statistically independent it follows that in the subpopulation of a given f the x_i are independent and therefore the y_i are independent. For example,

$$(8)\quad \Pr\{ y_1 = 1, \ldots, y_p = 1 \mid f \} = \pi_1(f) \ldots \pi_p(f) = \left[1 - \Phi\left(\frac{\alpha_1 - \lambda_1 f}{\sigma_1}\right)\right] \ldots \left[1 - \Phi\left(\frac{\alpha_p - \lambda_p f}{\sigma_p}\right)\right].$$

The case of one latent attribute (that is, one factor score) has been of particular interest to sociologists and social psychologists. The items may have to do with opinions on various questions, all of which are related to a given attitude. For instance, the items may be questions such as "Would you like to work with a negro?" "Would you like to live in a nonsegregated community?", etc. The underlying attitude is that of racial prejudice. The end purpose of the investigation may be to rate the respondents on a scale of racial prejudice.

In (7) $\pi_i(f)$ is given as a particular function of f. Other functions may be postulated. Lazarsfeld has been studying mathematical and statistical problems that arise when $\pi_i(f) = a_i + b_i f + c_i f^2$ or $\pi_i(f) = a_i + b_i f^{c_i}$ $(0 \leq f \leq 1)$.

Perhaps the simplest of the latent structure models arises when the latent attribute is discrete. Suppose that the attribute can have one of q values, say $f = 1, \ldots, q$. Then $\pi_i(f)$ can be designated as π_i^{α}, where $\alpha = 1, \ldots, q$. The distribution of the latent attribute is specified by $\Pr\{ f = \alpha \} = \nu_\alpha$. The probability of drawing an individual from the α-th latent class is ν_α; the probability of drawing such an individual and getting a positive response on the i-th item is $\nu_\alpha \pi_i^{\alpha}$; and the probability of drawing some individual and getting a positive response on the i-th item is

$$(9) \qquad \pi_i = \sum_{\alpha=1}^{q} \pi_i^{\alpha} \nu_{\alpha}.$$

Similarly, the over-all probability of getting positive responses on items i and j is

$$(10) \qquad \pi_{ij} = \sum_{\alpha=1}^{q} \pi_i^{\alpha} \pi_j^{\alpha} \nu_{\alpha}, \qquad i \neq j,$$

and the probability of getting positive responses on i, j, and k is

$$(11) \qquad \pi_{ijk} = \sum_{\alpha=1}^{q} \pi_i^{\alpha} \pi_j^{\alpha} \pi_k^{\alpha} \nu_{\alpha}, \qquad i \neq j,\ j \neq k,\ i \neq k.$$

There are similar expressions for the probabilities of other sets of responses.

The π_i, π_{ij}, and π_{ijk} are known or are estimated from the data; the π_i^{α} and ν_{α} are to be inferred. When the structure is known a respondent can be classified into one of the q latent classes on the basis of his responses to the items.

A method for inferring π_i^{α} and ν_{α}, suggested by Bert Green [2], is based on factor analysis methods. Equation (10) can be written

$$(12) \qquad \pi_{ij} = \sum_{\alpha=1}^{q} (\pi_i^{\alpha} \sqrt{\nu_{\alpha}})(\pi_j^{\alpha} \sqrt{\nu_{\alpha}});$$

these are identical with (3) if $\lambda_{i\alpha} = \pi_i^{\alpha} \sqrt{\nu_{\alpha}}$. The matrix (π_{ij}) can be factored to determine the matrix $(\lambda_{i\alpha}) = (\pi_i^{\alpha} \sqrt{\nu_{\alpha}})$, but there is left the indeterminacy of rotation. To eliminate this, Green suggests also factoring $\Sigma_k \pi_{ijk}$. Generally there will be only one solution that factors both matrices; in the case of fallible or statistical data there will be one factor matrix that gives the best fit. (This description of the method should be taken only as approximate.)

An important difficulty with this method is that π_{ii}, π_{iij}, etc. are not defined and not observed and these must be approximated in some fashion; this causes particular trouble with the second matrix $(\Sigma_k \pi_{ijk})$ that is factored. Another method [1] which bypasses this difficulty as well as the need for factoring large matrices is to use only part of equations (9), (10), and (11). The part of (10) that is used is a set that is entirely below the main diagonal of the matrix (π_{ij}). In this method $2q^2$ of the observed π_i, π_{ij} and π_{ijk} are used to infer the same number of the latent parameters. The algebra that is involved is the standard theory of characteristic roots and vectors, but we cannot go into details here. This method has been developed further to use more of the data, and a large sample theory of tests and confidence regions has been derived. Neither of the methods mentioned is in principle efficient in using all of the information in the data, but either one makes it possible to analyse data in terms of this model. However, further developments in statistical inference are needed even for the simple case.

REFERENCES

1. Anderson, T. W. On estimation of parameters in latent structure analysis. *Psychometrika*, 1954, 19, 1-10.
2. Green, Bert, Jr. A general solution for the latent class model of latent structure analysis. *Psychometrika*, 1951, 16, 151-166.
3. Lazarsfeld, P. F. A conceptual introduction to latent structure analysis. In P. F. Lazarsfeld (Ed.), *Mathematical thinking in the social sciences*. The Free Press, 1954, Pp. 349-387.
4. Stouffer, Samuel, *et al. Measurement and prediction*. Vol. IV. *Studies in social psychology in World War II*. Princeton Univer. Press, 1950. Ch. 10 and 11.

The "Moderator Variable" as a Useful Tool in Prediction

DAVID R. SAUNDERS

This paper is intended to be partly informative and partly persuasive. From the information standpoint, I hope to supply you with answers to three main questions: What is a moderator variable? How do you use a moderator variable? And, thirdly, Why should you, anyway? The persuasive aspect of the paper is obvious in the third question, just as in the title. It is an important aspect, because much of the information I can give you is neither very new nor very complicated.

Moderator variables have been used for many years by our friends in economics and agriculture to help in fitting regression surfaces to their data. In these applications the moderator effect is typically just one of many that are possible with a multivariate curvilinear regression based on a polynomial expansion. Our economic and agricultural friends don't need and don't seem to have any special name for it. Our biological friends might be tempted to suggest the name "synergistic variable," but this is a term that already has a lot of additional scientific connotations that we want to remain neutral on.

So far as I know, Gaylord and Carroll were the first to use a moderator variable in psychology. They called it a "population control variable," and presented a paper on it during the 1948 APA meetings. The term "population control variable" is a good one, because it suggests a very important application. But it is a bad term to the extent that it tends to blind us to a number of other equally important applications, which I will touch on. The term "moderator variable" seems to be general enough in its meaning, and still not to be loaded with too many undesirable connotations.

Well—we have just christened this thing, but have only hinted at what it is. Let's look at some examples. By now I've managed to think of dozens of attractive hypothetical examples, but I'll spare you these and emphasize two examples that have been fully worked out and cross-validated.

The first example is the one that originally led me to think a moderator variable might be an important concept. Frederiksen and Melville, at ETS, had just shown that interests were less predictive of academic success for "compulsive" people than for "non-compulsive" people (2).

For this discussion, we can ignore the measures that were used to define interests, success, and compulsiveness, except to note that they *all* were regarded as continuous variables. Frederiksen's and Melville's experimental design, nevertheless, had to be the analysis of covariance; the total experimental sample was arbitrarily divided near the median compulsiveness score to produce two groups—called "compulsives" and "non-compulsives." The relations between interest and success were compared for these groups, and were found to differ significantly in the slopes of the regression lines. This was done separately for ten different interests as predictors.

This example is typical of many situations in which the use of a moderator variable should be considered. In this particular kind of situation the term "population control variable" would also be apt. Clearly, what we are after is a means of treating compulsiveness as the continuous variable which it is—a means of avoiding the arbitrariness of dichotomizing or otherwise dividing the population into *smaller* pieces—a means of maintaining the integrity of the total population while still maintaining a statistical control on each individual's membership in one of a continuous, infinite series of sub-populations defined by his compulsiveness score. In short, we will allow the meaning of his interest score to be "moderated" by his compulsiveness score.

Mathematically, this turns out to be extremely simple to do. Suppose we start with an ordinary linear regression using several variables. Keeping everything in standard scores, we would write the equation

$$y = \sum_{i=1}^{k} \beta_i x_i,$$

where y is our criterion, and β_i is the beta-weight for predictor x_i. Now suppose that β_i, instead of being a constant, is itself a linear function of a series of moderator variables, z_j's. If we plug this into our equation, and do a little rearranging, we immediately find that we can write

$$y = \sum_{i=1}^{k} a_i x_i + \sum_{j=1}^{l} b_j z_j + \sum_{ij} c_{ij} x_i z_j.$$

This would be just another linear regression if it were not for the last term, involving the products of the x's and the z's. If we want to, we can always choose origins of measurement for the x's and z's that will make everything drop out of this equation *except* the products, and a constant term.

It is evidently these product terms which are inextricably tied up with the use of a moderator variable. And that is all that has happened. So long as there is a clear separation of the x's and z's, we cannot have any use for squared variables, let alone terms of higher power. All we have to do to fit the model to data is to find the appropriate xz product (or products) for each subject, and treat it (or them) as new independent predictors in any standard multiple correlation technique. Of course, we cannot introduce a product variable into a battery unless both of its factors are already there.

There are many interesting mathematical sidelights to this thing, but if I want to persuade you that moderator variables are useful and practical, we'd better go back to our examples.

We took the data that Frederiksen and Melville had collected, and computed the products of the interest predictors with compulsiveness for each individual subject. Then we ran multiple correlations, first adding compulsiveness to the interest predictor, and then adding their product to the battery (3).

For three of the ten interest predictors studied, the simple addition of compulsiveness to the battery gave a significant increase to the multiple R. In these three cases the compulsiveness score happened to act as a suppressor variable. A moderator variable is different from a suppressor variable, though they both typically have zero zero-order relation to the criterion; a moderator variable does not have to have zero-order correlations with even the predictor.

Back to the example: In five of the remaining seven instances, addition of the appropriate product score to each battery of two measures resulted in a further significant increase in the multiple R. The sign of the beta weight for the third term was correctly predicted from the hypothesis in all ten of the ten instances. You will recall that the hypothesis told us at which end of the compulsiveness scale to look for good predictions.

These results looked promising. So we moved the scene of operations from Princeton to Rochester; from a group of self-referred counselees to a larger group tested routinely during Freshman Week; from Strong Interest Blanks scored with weighted responses to scoring with unit weights. The criterion was still freshman grade average for engineering majors, and the moderator was still the Accountant Interest score of the Strong, as a measure of compulsiveness. Insofar as the same or similar interest scores were available to try as predictors, we were able to cross-validate all but one of the statistically significant findings from Princeton.

In this example we have observed that in the more significant instances, the predictive contribution of the moderator effect is just as great as the initial contribution of the interest variable predictor. This may not seem like saying much, but note that Frederiksen and Melville reported correlations ranging all the way from zero to over fifty for sub-groups of the same population, divided using a pair of compulsiveness measures. Even when a moderated regression model does not lead to much increase in the multiple R, it does lead to quite different predictions for some individuals, and should be used if its effect is even statistically significant. These predictions may differ in standard error as well as in the expected value itself.

In the second main example that I want to discuss here, *neither* the moderator *nor* the predictor has been shown to have a significant zero-order correlation with the criterion! In this situation, the predictor and moderator lose their separate identity, and a name like "population control variable" becomes more awkward to use. In this example, there is no significant multiple R until the product term is introduced; then it jumps to values like .45. This example is based on some of Fiedler's recently reported work on the influence of leader-keyman relations on small group effectiveness. Many of you may have seen this written up in *Time* magazine recently, even if you missed Fiedler's APA presentation and haven't yet seen his report to the Office of Naval Research (1).

The gist of it is this. Suppose you are the formally designated leader of a small group or team. There is probably someone in the group who is your right-hand man—your principal subordinate or keyman. If you are the kind of person who generally has warm feelings for most people, for whom similarities among other people are more important than differences, it will pay you and your group for you to maintain a relative aloofness from your group, and especially from your keyman. On the other hand, if you tend to think of others you like as being different from those you dislike, it will pay to do the opposite, namely for you to cultivate strong sociometric ties with your group, and especially with your keyman.

Here, then, are two variables you need to measure to predict a particular leader's effectiveness in a particular group. While neither variable is related to the criterion, their product has a substantial negative correlation with it. The results happen to be psychologically very sensible, and they can probably be used to counsel leaders towards a more effective style of leadership, and to predict what groups will respond best to given styles of leadership.

These two examples have been very different in many respects, but they do have two things in common in addition to their featuring of important moderator effects. For one thing, they both feature non-cognitive variables as predictors—and as moderators, if you still care which is which. It seems to me that we can expect to find examples of this kind relatively more easily, but cognitive examples are still a real possibility. For instance, if we could obtain a meaningful coefficient of reliability for an individual's test score, independently of the group within which he happened to be tested, it would probably moderate any predictions made from the test score itself.

In the second place, both these examples were initially studied by breaking a total sample up into sub-groups. There are many studies that have been carried this far, and then reported in tedious detail for lack of an organizing concept such as the moderator variable provides. For the past three years I have made it a point at the APA meetings to seek out studies of this kind; I have always found my fill without going very far afield. For some reason, studies involving the "Authoritarian Personality Syndrome" seem to be in particular need of something like a moderator variable. But no more so, I would say, than people who do configural scoring or make clinical judgments on the basis of personality tests.

There is one last topic on my agenda. Assuming you have decided to look for moderator effects, what is the best way to go about it? There are at least four methods to consider. 1. You can start with a good hypothesis. This is always a good idea. 2. If you don't have one handy, you can look for one by studying a few cases intensively, and seeking out variables whose interpretation seems to depend on other variables. 3. You can do the same thing with larger groups of cases by looking for sets of sub-groups within which correlations are significantly different. 4. You can do the same thing with items instead of variables, by testing the interaction variance of a pair of items against a criterion. This fourth approach is a whole technique in itself, and I wish there were time to tell you about it. It capitalizes on the electronic computers; it can be generalized beyond pairs of items; and it brings the old idea of keying patterns of response to several items into a new perspective.

REFERENCES

1. Fiedler, Fred E. The influences of leader-keyman relations on combat crew effectiveness. Urbana: Group Effectiveness Research Laboratory, University of Illinois, June 1954. (Technical Report No. 9, Contract No. N6-ori-07135.)

2. Frederiksen, N., and Melville, S. D. Differential predictability in the use of test scores. *Educ. psychol. Measmt.*, 14, in press.

3. Saunders, D. R. Moderator variables in prediction. Princeton: ETS Research Bulletin 53-23, October 1953.

A Method of Factoring Without Communalities

L. L. THURSTONE

Ever since multiple factor analysis was developed over twenty years ago, the unknown diagonal cells of the correlation matrix have been a serious problem. These are called the communalities and they represent that part of the total variance of a test which it shares with one or more of the other tests in the battery. For each method of factoring it has been necessary to estimate these diagonal values in the correlation matrix and various methods have been devised for doing so.

In a theoretical case one can construct a correlation matrix of order n and rank r where $r < n$. The diagonal entries of such a correlation matrix are then so determined that all minors of order $(r + 1)$ vanish. When the order n is larger than, say, 6 or 8, the diagonal values are usually unique. In dealing with experimentally observed correlation coefficients, the coefficients necessarily have variable errors so that a low rank cannot be found in any exact manner. However, one can usually write another correlation matrix with side entries that are nearly the same as the experimentally observed values and which is of a rank much lower than n so that, for example, $r < n/2$. This is the situation often found in multiple factor studies.

Various methods of factoring the correlation matrix have been devised in which the diagonal elements are first estimated. One of the simplest is to assign to each diagonal cell a value equal to the absolute value of the highest entry in the corresponding column or row. The estimate is revised after each factor has been extracted. This rough method of estimation is quite successful when the order is fairly large, say, about 15 or 20 or higher. Although this procedure is quite successful for many scientific problems with correlation matrices of high order, it is far from satisfactory for theoretical formulations of the factoring problem.

There is an important relation between the communalities and the number of factors that are used for describing a correlation matrix. The larger the number of factors, the higher are the communalities. The communality of any test variable j is the sum of squares in the corresponding row of the factor matrix F. If it is decided that the first r principal axes are inadequate for the description of the correlation matrix, then the next principal axis may be determined. This gives (r + 1) columns of the factor matrix. The communality of any test j is then augmented by the square of the entry in column (r + 1) in the row j. But it is never known beforehand how many principal axes need to be determined in order to reduce the residual correlation coefficients to values that can be ignored in terms of their sampling errors. Hence we have the anomalous situation of not knowing at the start of the computations how many factors need to be postulated to account for the correlations. The communality estimate should rise with the number of factors that are extracted but the relation depends entirely on the unknown configuration of test vectors in each given problem.

In practice the problem is resolved by adjusting the communalities for each factor that is extracted from the correlations or by repeating the whole factoring process until the estimated communalities at the start agree with the sum of squares of the factor loadings in the rows of the resulting factor matrix. But this process depends on a certain number of factors as determined by the first cycle. The adjustment should be done over again if the investigator decides to increase the number of factors used. When the number of factors is quite large, such as 10 or 15, then the adjustments in the diagonals are ordinarily quite small for each additional factor.

In trying to relate the theory of multiple factor analysis to practical scientific work with large correlation matrices, this situation is evidently quite unsatisfactory, even though the practical compromises have been adequate to resolve most of the scientific problems so far in the isolation of the components of human intelligence which have been called primary mental abilities.

Several months ago I was sitting in an airplane in Helsinki in Finland, waiting for the take-off for Stockholm. It occurred to me suddenly that this awkward situation that we have fought and compromised with for a long time could be resolved in a ridiculously simple way. I shall describe the idea here although we have not yet developed the best computing methods, and I hope that some other students of the multiple factor problem will consider the new method and how it might be reduced to the simplest possible computing procedures.

In the experimentally given correlation matrix, the diagonal cells are unknown. They are certainly not experimentally given. Let us determine the first column of the factor matrix by the method of least squares so that the first factor residuals are a minimum. This is precisely what we do in determining the first principal component by Hotelling's iterative method *with one important exception*. This exception is that we ignore the unknown diagonal cells completely. Since they are unknown, they are not represented in any observation equations. Only the experimentally known correlation coefficients participate in the observation equations for the least squares solution. When the first column of the factor matrix has been determined so as to minimize the residuals for the side correlations, then we find the first factor residuals for the side entries. The diagonal cells are ignored completely. Then we proceed as before for each additional column of the factor matrix until the residuals of the side correlations can be ignored. The communalities for this given number of factors are then simply the sums of squares of the rows of the factor matrix *F*. In this procedure we determine the communalities after the factoring job has been completed. If, for any reason, we decide to reduce the residuals still further, that can be done by extracting another factor. The new communalities are then merely the sums of the squares of the rows of *F* with another column added. None of the factoring need be repeated because communality estimates are not involved in the factoring of the correlation matrix.

It was several weeks until I had the opportunity to try the new method in Frankfurt, Germany, where I discussed this method with three of my former students in Chicago. They were Dr. Hans Anger, Dr. Sten Henrysson, and Rolf Bargmann. Bargmann set up three test cases and reported that the solution gets close to the principal axes solution and Dr. Sten Henrysson has reported similar findings with the method in Uppsala. When I return to my laboratory in Chapel Hill, I expect to investigate the method further.

The computational procedure can be tried in a manner analogous to Hotelling's iterative solution but it is likely that one of several other alternatives will be more effective. In one manner of writing the problem we get third degree normal equations which can be solved by successive approximations with additive corrections to the assumed factor loadings.

There may be an interesting geometric twist to this problem in that the customary geometric model may have to be revised but I am not prepared yet to elaborate on the geometric implications of this problem. The slight disturbance of the geometrical model for the correlation coefficient and the factor matrix can be seen with a theoretical case of exact rank r. When this new method is used, the r th factor residuals do not vanish identically. They are small, however. The factoring matrix obtained can still be given the usual geometrical interpretation and the correlation coefficients are closely represented as scalar products of the test vectors that are defined by the factor matrix. It is a question to be investigated whether the disturbance of the true geometrical model is less important than the disturbance implied in all the adjustments of the communalities in practical computing procedures.

This procedure calls to mind Spearman's formula for the best fitting single factor. However, the application of such a formula may lead to trouble if applied to the residuals for subsequent factors after the first factor. The theory of this problem should also be investigated with reference to the possible appearance of the Heywood case. So far we have not encountered it. In several trials we find factorial solutions that are close to the principal axes.

We hope that other students of factorial analysis may be interested to explore this factoring method.

DISCUSSION

PARTICIPANTS

T. W. ANDERSON, EDWARD E. CURETON, FREDERIC LORD, FRANK ROSENBLATT, P. J. RULON, DAVID R. SAUNDERS, ROBERT L. THORNDIKE, L. L. THURSTONE, LEDYARD R TUCKER, JOHN W. TUKEY, JOSEPH ZUBIN

DR. ZUBIN: Dr. Anderson, did you make each probability of response independent, or would you allow for interaction, for instance, between saying yes to one item and no to another? Is the probability of that pattern based on the product of the original probabilities, or might there be some interaction?

DR. ANDERSON: The assumption here is that given the latent structure score, then the responses are independent. If one estimates the latent parameters, and then writes down the right-hand sides of the equations and finds they do not correspond so well to the left-hand sides, that would indicate that the assumption is not borne out very well. Then one can extend the model so that essentially, instead of basing this on one latent score, you have several latent scores. For a particular couplet, let us say, you may have to throw in an extra latent score to account for the kind of interaction you are talking about.

DR. RULON: Are there any other questions? I do not know whether Dr. Anderson made himself so clear that there are no questions, or so obscure that there are no questions.

DR. CURETON: I have a comment that might have a bearing on Dr. Zubin's question. It is to be noted that the number of latent classes is so determined that the responses will be independent within such classes as far as that can be determined statistically. In other words, you simply postulate enough latent classes so that you do have independence within each one of them.

DR. ZUBIN: Is there any way of telling how many subgroups you have in the overall sample?

DR. ANDERSON: One technique for doing that is to carry over what you do in factor analysis, again using the analogy between these equations. In factor analysis, there are techniques to determine how many factors there are, and the same techniques can be used here to determine how many latent classes there are. However, from the point of view of the theoretical statistician, these techniques are not as well worked out, or they do not apply as well, as in the ordinary factor analysis.

Dr. Cureton: Dr. Saunders, at the same time that you put in the "moderator variable," the formal product, wouldn't it be equally easy to put in the squares of the X's and Z's, so that the direct regressions are stretched from linear to parabolic?

Dr. Saunders: In general, there is no reason why this cannot be done. Again, in general, if you do do this, you are increasing the number of parameters that are going to be determined from the experimental data. In the absence of some kind of evidence to suggest that these square terms are important, I think it would be a cautious approach to steer clear of them. Let us be open-minded, to be sure.

Dr. Thorndike: Along the same line, isn't a fifth possible way of exploring these moderator variables just routinely to compute all the products and run them in as additional variables in your basic prediction enterprise?

Dr. Saunders: This is an approach, and it is a lot of work.

Dr. Cureton: You were talking about electronic computers.

Dr. Rosenblatt: If you start with a matrix of variables, you could compute an ordinary linear multiple r for this matrix, and take the coefficient of determination (or r^2). If you compare this coefficient of determination with that of a correlation which includes the product terms, have you given any thought to how the difference between these correlations stands with respect to the usual notion of statistical interaction, coming out of an analysis of variance? It seems to me that the correspondence becomes quite close (i.e., the incremental correlation due to the moderator variable corresponds closely to the interaction variance).

Dr. Saunders: I would say that it is quite close. Incidentally, this spells out an approach that you can take in testing the significance of the product term. You need to compare these coefficients of determination with and without the product terms.

Dr. Thurstone: This is exactly the same thing as we get in Hotelling's procedure, except that the sum of the squares of the factor loadings in Hotelling's method is the sum of the squares in all of the factor loadings and the corresponding relations here would include the diagonals. In this case we deal only with the known values. But if you write zero in the diagonals, then this term can be interpreted in an ordinary way. This is the sum of the squares of all but one of the factor loadings.

One can proceed in this manner, which is very closely analogous to Hotelling's procedure.

We have found that the convergence in several problems is fairly rapid. If the convergence should turn out to be slow here there is no sense in carrying out fifteen or twenty or more iterations. If the next two dimensions are nearly circular, so that the next two latent roots are nearly the same, then the convergence is known to be slow. When that happens, we should be practical enough to realize that it really does not make any difference where you put the next axis. You can stop after about the third iteration. If you are talking to a mathematician, he might not accept it, but for practical work, it is useful.

Dr. Tukey: I should like to raise several specific questions. One: if we are going to do this on the least square basis, what is the philosophy that guides us to use unweighted squares of deviation? If we were looking solely at sampling variations, which we are not, then we would want to weight this quite a lot differently for large correlations than for small. For the purpose of factor analysis as a whole, I am not clear how we should weight it. I am just asking for guidance.

Dr. Thurstone: You are not referring to the A's as weights.

Dr. Tukey: No, I am referring to the possible weights outside the parenthesis.

The second question I should like to raise is this: if we are going to give up the communalities and go ahead in this direction, putting aside the difficulties of mathematics in computation which might be really serious, if we decide to take out two factors, why should we take out for the first one, the one that comes first and singly by this method? Would we not be better off, in principle, maybe not in practice, to accept this formula and minimize with respect to the A's and the B's at the same time? If one could do this, in this example you cited of Rank R, you would get everything if you could do it for all R's together.

Dr. Thurstone: Yes, I would rather write the equation with R plus one or two or three terms. You could then discard some if you want to. However, suppose you gamble that there are, say, ten factors. You write ten terms. It gets a little unwieldy to handle. That is why we do not do it.

Dr. Tucker: I have one question about the examples you have tried. Did some examples have somewhat equal communalities while other examples had quite variant communalities? I suspect the range of communalities makes quite a bit of difference in the closeness of approximation by your method. If the communalities were to come out the same, your method would yield an exact solution.

Dr. Thurstone: That is a good point to make. As a matter of fact, we set up several empirical cases. For small variation in communalities, the convergence was much more rapid. In a case with wide difference of communalities the convergence was slower. I think you are quite right in that respect.

Dr. Lord: You pointed out the procedure was equivalent to putting zeroes in the diagonals. What is the advantage in having zeroes in the diagonals rather than ones? Is it an approximation to something else?

Dr. Thurstone: That is not the same problem. You are raising the question of why we should use communality instead of unity in the diagonals.

Dr. Lord: No. If you are not going to use the communalities, are you better off using zeroes rather than ones?

Dr. Thurstone: My objective is to ignore the diagonals, and they do not participate in the equation. That is the principle of the method. If the diagonals are unknown, let us leave them blank. We do not have to write zeroes—just leave them blank, they do not participate. However the computations would be the same as if the diagonal correlations were zero.

SECTION 2

Evaluating Group Interaction

A New Technique for Measuring Individual Differences in Conformity to Group Judgment

RICHARD S. CRUTCHFIELD

Central to research both in personality and in group dynamics are methods for the measurement of the individual's behavior in situations of group-interaction, for example, conformity behavior to group pressure.

Such measurement is made difficult by certain demanding requirements. (1) Ideally, the behavior should be measured *directly* in actual group situations, rather than indirectly by questionnaires about group situations. (2) The group situations should be psychologically *relevant* for the individual. (3) There should be *standardization* of the group situation so that measurements of different individuals may properly be compared. And (4) there should be adequate *economy* of the test method, so that substantial numbers of persons may be tested without unreasonable cost in time and money.

The standardization problem is the most acute. In genuine groups, involving face-to-face interaction of several subjects, the stimulus situation confronting each person is *unique*, being dependent in part upon what the others in the group do. This leads to an undesirable confounding of personal and situational factors in the measurement of the individual's group behavior, and there is no simple way to disentangle the two.

One fruitful attack on this problem is what in previous research applications I have called the "quasi group-interaction method." Its essence is simple. Subjects to be measured are placed together in a group situation which, though perceived as genuine by them, is actually so contrived by the experimenter that he wholly controls and manipulates the conditions of group-interaction. This serves to standardize the situation for each subject in an identical fashion, so that observed individual differences in behavior may properly be ascribed to differences among the persons rather than to situational differences.

The quasi group methodology is by no means new. Earlier variants are found throughout the field of experimental social psychology.

Little, however, has been made of it in personality measurement, and in general the method has not been widely exploited.

In recent work I have applied it to what gives promise of being a powerful new technique for measuring individual differences in conformity to group judgment. Five persons are tested simultaneously. They are seated in a row in front of an electrical switchboard apparatus, consisting of five identical panels. Each panel is shielded from the other four. It contains eleven numbered switches by which the person may signal his response. It also contains five rows of signal lights which can display the responses being made by the other four group members. In short, there is a simple electrical communication system among the five persons. No *direct* communication is permitted.

The task for each person is to make judgments pertaining to a large number of slides which are projected on a wall facing the group. Each slide offers a set of numbered alternative answers among which he is to choose. He records his choice by closing the appropriately numbered switch and this causes his response to be displayed on the panels of the other four members. He is also instructed to respond in a specified serial *order* within the group, person A going first, then B, and so on, E being last. The designation of his letter position—A, B, C, D or E—is indicated to him by the experimenter. Such letter designations are rotated from time to time, permitting each person to respond in each of the five serial positions. The experimenter offers no further explanation of the purpose of this procedure.

The slides offer a mixture of materials—simple *perceptual* comparisons, such as of relative length of lines; *logical* problems, such as the completion of a number series of the kind found in standard mental tests; expression of one's own *opinions* and *attitudes* on various issues, etc.

On the first set of slides the person finds the judgments fairly easy to make, and he observes that there is a sensible agreement between his judgments and those of the other four members. But when later in the series he is for the first time designated as E, so that he must respond in last position, he experiences something new and disturbing.

On this slide, calling for a simple judgment of relative length of lines, he sees the other four members unanimously agree on a choice which clearly contradicts his own. This throws him into a severe conflict between the clear evidence of his own senses and the unanimous contradictory consensus of the rest of the group. How he chooses to resolve this conflict, either by yielding to the group pressure and conforming to its judgment or by remaining independent of it, is the basic measure of conformity behavior in our procedure.

There is, of course, more than one such critical slide. While responding in position E, he is presented with more than twenty other critical slides—pertaining variously to matters of perception, of logic, of opinion, and of attitude—on each of which he is confronted with a serious disagreement between his own judgment and that of the rest of the group.

As you will doubtless have surmised, the situation is not really what the persons have been led to understand. They have been grossly deceived. The five panels are not connected to one another at all, but to a control board operated by the experimenter. It is he who signals the responses which allegedly come from the other group members. The wiring is in parallel so that the pattern of signals he chooses to simulate appears simultaneously and identically on all five panels. Moreover, the designations of serial order of responding—A through E—are likewise identical for all five persons at every moment.

By this deception, therefore, a quasi group-interaction situation has been contrived, permitting each person to be exposed to a standardized set of simulated group judgments and to be confronted at predetermined points with identical external pressures toward conformity.

Since the Spring of 1953 when this technique was first developed, three studies have been made using different populations of subjects. The first study was of 50 men, averaging 34 years of age, all members of a profession in which leadership is one of the most important qualifications. They were tested at the Institute of Personality Assessment and Research as part of a larger assessment program. The second study was of 59 college students, mostly sophomores. The third was of 50 women, all college alumnae in their early forties being studied under the auspices of the Mary Conover Mellon Foundation.

The results of these studies, which attest to the technical success of the method and throw light on the nature and determinants of conformity behavior, are summarized with great brevity in the following eight points.

1. *The general amount of conformity behavior exhibited is large.* A single fairly representative item will serve to illustrate. A circle and a star are exposed side by side, the circle being about one-third larger in area than the star. The false group consensus is on *star* as the larger, and as a consequence 46% of the men give this same false judgment.

2. *The degree of conformity shown depends in part on the kind of material being judged.* Although some amount of conformity can be elicited on every one of the critical items, the range in effectiveness among them is extremely wide. At the lower end, a simple expression of *personal preference* for one of two line-drawings is very little suscepti-

ble to a contradictory group consensus, the degree of conformity varying from zero to ten percent in the several studies. At the upper end, the most effective item is one which has been deliberately maximized for ambiguity. The subjects are asked to complete a number series, for which there is actually no logical solution. When the simulated group judgment agrees unanimously on an obviously illogical completion, 79% of the men express agreement with this answer.

3. *Substantial conformity is elicited on socially important judgments as well as on more abstract judgments.* A critical methodological issue is whether the conformity effects so far mentioned merely represent rather superficial readinesses of the person to agree with the group on matters of little real importance to him, or whether instead they do reflect more basic conformity tendencies in the person. Support for the latter interpretation is found by the introduction of new critical items in the studies of college students and mature women. These items called for expression of the person's attitudes on matters of civil liberties, subversion, ethics, crime and punishment, etc. Pronounced conformity effects are found on these socially and psychologically relevant items. Take a single example from the study of college students. The question was asked, "Which one of the following do you feel is the most important problem facing our country today?", and these five alternatives were offered: economic recession, educational facilities, subversive activities, mental health, crime and corruption. Among control subjects tested alone, only 12% chose "subversive activities" as the most important. But when exposed to a spurious group consensus unanimously making this choice, 48% of the subjects gave this answer.

4. *There are pronounced individual differences in amount of conformity shown.* A total conformity score for each person may be readily computed by summing the number of the critical items on which he exhibits conformity to the false group consensus. Virtually the entire possible range of such scores is found in all three studies. Among the men, for instance, at the lower end several subjects showed practically no conformity, being influenced on one or two items at the most. At the upper end one man was influenced on 17 of 21 critical items. The scores are well distributed between these extremes, with a mean score of about 8 items and a tendency for greater concentration of scores toward the lower conformity end. As estimated from the correlation of sub-scores on two matched halves of the critical items, the reliability of the total score for the sample of men is .90.

5. *There are both generality and specificity in the conformity tendencies among individuals.* Although there is a generally positive matrix of intercorrelations of degree of conformity effect for the various items,

there are also some useful differentiations to be made among the items. A cluster analysis yields one principal set pertaining to highly structured items, i.e., those involving clear judgments and unambiguous stimuli. A second main cluster consists of poorly structured items, i.e., those involving uncertain judgments and ambiguous stimuli. The individuals can be scored separately on these clusters. There is evidence that these cluster scores as well as individual performance on single items must be taken into account in the analysis of conformity behavior.

6. *The degree of conformity behavior relates significantly to relevant dimensions of personality.* Validity of the conformity measure is attested to by its substantial relationships with numerous ratings, objective test scores, and other personality determinations in assessment of the sample of men. Those low in conformity, that is, those who successfully resist the group pressure, can be clearly characterized as having intellectual effectiveness, ego-strength, self-acceptance, leadership ability and maturity of social relations. The high conformists, on the contrary, reveal inferiority feelings, rigid and excessive self-control, intolerance of own impulses and lack of self-insight, authoritarian outlook, emphasis on external and socially approved values, and disturbed attitudes toward other people. This general picture coincides well with prior theoretical and empirical studies of the personal traits associated with conformity, but, of course, it represents a gross oversimplification of the complex relationships yet to be explored.

7. *There are significant differences in average amount of conformity exhibited by the several populations of subjects.* The college student sample was made up of both males and females. The male students exhibited just about the same average level of conformity as previously found in the adult, professional men. But the female students exhibited significantly *more* conformity than these male groups. On the other hand, the sample of mature women—college alumnae in their early forties—showed significantly *less* conformity on the average than that found in all the other groups.

8. *Experimental variations in the testing situation can change the amount of conformity shown.* We have seen that there are differences in degree of conformity relating to personality determinants, to the nature of the populations of subjects, and to the kind of items being judged. We can also show differences resulting from certain experimental changes in the situation itself. As one example, when a group of student subjects were exposed to additional pressure, namely, that given by having the experimenter later confirm the "correctness" of the false group consensus, conformity effects were markedly increased.

To revert now to our initial statement of requirements for proper measures of individual behavior in situations of group-interaction, namely, directness, relevance, standardization and economy, it would appear that the present method fulfills all these requirements. The very essence of the method is such as to guarantee standardization of conditions. It is clearly economical, providing for the testing of five persons at once in a period of approximately one hour. It is direct, in that the persons perform in a situation which is for them one of real group-interaction. On this point it should be emphasized that the genuineness of the situation is virtually never challenged by the subjects. Of the total of 159 persons already tested in the procedure, only a small handful when questioned immediately afterwards expressed doubts of its genuineness; of these only two or three seemed to have felt such doubts while actually in the situation.

Finally, as to psychological relevance, we may judge from both the manifest behavior and the retrospective reports by the subjects that the situation was deeply involving and that the group pressure created anxiety, often acute. A substantial number of persons freely admit on later questioning that they had violated their own inner convictions in order to give responses in accord with those of the group.

The findings attest to a rich potential use of this measurement technique when applied to a variety of problems—personality assessment, group dynamics research, sociological and cross-cultural comparisons of groups, experimental and theoretical study of the psychological conditions of conformity. A number of such studies are contemplated and several are now in progress.

The Russell Sage Social Relations Test: A Measure of Group Problem-Solving Skills in Elementary School Children

DORA E. DAMRIN

The test I am going to describe to you this morning represents an attempt that has been made by myself and certain of my colleagues in ETS to develop an instrument of measurement that will get at certain of the intangibles in education. As all of you know, over the past several years there has been increasing emphasis, particularly in the field of elementary education, given to the idea that children should be taught to work together cooperatively, to share things with each other, to participate in the planning and carrying out of certain classroom activities, and to conduct themselves in ways compatible with a free democratic society.

The test I am going to show you today is designed to measure the extent to which some of these things have been accomplished. Since the test is a little unusual, I thought that it would be best if I would pretend that you are a group of students and give it to you in the same way that we give it to elementary school classes. The score sheets you have will be explained after this demonstration.

The examiner enters the classroom and sets up a card table in the front of the room. He then says, "Today, boys and girls, I am going to give you a test. Now, you know usually when you take a test, each one of you sits in his own seat, does his own work, and you are not allowed to help each other. Isn't that correct?"

The students usually nod rather solemnly in the affirmative at this point. The examiner goes on to say, "But this test I am going to give you today is exactly backwards. In this test you are *supposed* to help each other, you are *supposed* to work together, and you are *supposed* to talk to each other while you are taking it." Here the children usually evidence some amazement and seem to wonder what in the world is coming next.

The examiner then says, "Let me show you the first problem on the test so that you will understand what it is about." The examiner opens the testing case and holds up this rather attractive form, built of different colored plastic blocks, and says, "Now, what does this look like?"

The children respond with, "That's a house."

The examiner continues, "Yes, that's right, this is a house, and I am going to put it here on the card table for you to look at all during the time you are taking the test."

"Now here in this box, boys and girls, I have all of the blocks that you will need to build this house, and what I am now going to do is pass around the room and give each of you one or two of these blocks. Then I want all of you to get together and see how quickly and how well you can build this house."

The examiner goes around the room, giving each child one or two blocks, depending on the number in the class, and at that instant we start scoring the test, continuing on until the children have completed the problem.

After the blocks are passed out, the children are told, "Now before you start you can take all of the time you need to plan how you are going to get together and do this, but after you tell me that you are ready, then I am going to begin timing you, and you will be given fifteen minutes in which to build the house. If you complete your house *before* that time, your score on the test will be much higher."

The examiner then works with the children while they build a plan, and all during this time their comments and suggestions are scored. When the children indicate that they are ready to begin, the examiner scores them on how well they carry out their plan. At the end of the first test problem the examiner presents them with the second, and finally with the third problem.

The second and third problems are progressively more difficult than this one. If you would be interested in examining them, I will be glad to show them to you afterward, since our time is somewhat limited.

Unfortunately, in my discussion of the scoring of the test, I cannot present data on the validity or the reliability of the instrument. It is of too recent development to do that. We are, however, presently conducting a major validity study in cooperation with Dr. Orleans and his research staff in New York City.

On the basis of the data which will be collected within the next six months or so, I hope that I will be able to support some of the things that I am going to hypothesize about this morning.

The test has two separate and relatively independent parts. The first is called the *Planning Stage* and the second, the *Operations Stage*. The Planning Stage begins at the time the examiner starts passing out blocks to the children, and ends when the children inform the examiner that they are ready to begin constructing the model. The Operations

Stage begins when the examiner tells the children "Begin" and starts his stop watch, and ends when the 15 minute time period has elapsed or when the children finish building the model. A set of six scores is obtained for the Planning Stage and a set of three scores for the Operations Stage.

The first score given in the Planning Stage is termed *Group Participation.* It is based upon the four specific behaviors listed on your Observation Sheet at the top of the page. The Observation Sheet is checked as follows:

1. Individual: A check is entered each time an individual class member makes a problem-relevant comment.
2. Buzz Groups: A check is entered each time one-half or more of the class members spontaneously turn to each other in small groups and begin discussing various elements and aspects of the problem. Many children speak at one time, comments are not individually distinguishable, but it is clear that the children are engaging in problem-relevant remarks.
3. Chorus: A check is entered each time one-half or more of the class members spontaneously chorus their opinion about a previously made individual comment. This chorus may indicate either agreement or disagreement with the idea.
4. Noise: A check is entered each time one-half or more of the class members become noisy and undisciplined. Many children speak at one time, the examiner is unable to hear individual comments, but it is clear that the children are engaging in many problem-*irrelevant* remarks.

The pattern of checks made by a group of children over these four categories of behavior is translated into a numerical score which reflects the amount and quality of the children's total participation. A high score on this variable indicates that all members of the group participated actively and freely in the planning stage, whereas a low score indicates a minimum of free participation. It is our hypothesis that the Group Participation Score is closely related to the amount of freedom which exists in the classroom of a particular elementary school teacher.

The second score is *Communication Pattern.* It is concerned solely with the intent of the verbal comments made by the individual children who contribute to the planning discussion and who are checked in the Group Participation Score. All comments made by individual children are classified in one of the four categories listed second on your Observation Sheet. These categories are defined as follows:

1. Suggests discrete idea: A check is made if the idea is one which has not previously been stated and which does not incorporate any part of or build on a previously mentioned idea.
2. Repeats idea: A check is made if the comment is either a rephrasing or a restatement of a previously mentioned idea.
3. Evaluates idea: A check is made if the comment takes a previously mentioned idea and points out why it is wrong, why it won't work, or why it is a good suggestion and should be adopted.
4. Improves idea: A check is made if the comment takes a previously mentioned idea and builds on to it so that the end product is more comprehensive, more precise and more detailed than the initial idea.

The pattern of checks made by a group over these four kinds of statements is translated into a numerical score which reflects the way in which children converse with each other in building a mutually agreeable plan of action. A high score is obtained by the group whose members improve and build upon each other's ideas and progressively work toward a final plan that is more precise, more detailed, and more comprehensive than the initial idea for a plan with which the group started. A low score is obtained by the group whose members' comments are largely discrete and repetitive. In this score a series of discrete and repetitive checks indicates that children are competing with each other to give "different" ideas rather than cooperating with each other to build a constructive solution. Thus the significance of the Communication Pattern Score resides in the fact that it provides information about whether the members of a particular elementary school classroom function as a collection of single individuals or as a cooperative group.

The third score is termed *Organizational Efficiency* and is based upon the content of the ideas which children propose during the planning period. Fortunately the major ideas which children can have about ways of solving the test problems are finite, and all possible suggestions can be classified into one of the ten categories listed on the third section of your Observation Sheet. These categories have the following definitions:

1. Construction Authority: Having one or a few children perform all of the actual building
2. Supervisory Authority: Having one or a few children supervise and help others who are building the puzzle
3. Simultaneous Groups: Organizing the class into groups on some basis and having all groups working simultaneously

4. Sequential Groups: Organizing the class into groups on some basis and specifying the order in which the groups are to take turns in building the puzzle
5. Problem Groups: Organizing the class into groups on the basis of some structural element of the puzzle, e.g.:
    a. By colors: red, white, blue
    b. By parts: neck, body, legs, tail
    c. By shape of block: square, triangle
    d. Any combination of the above
6. Non-problem Groups: Organizing the class into groups on the basis of something *external to* the puzzle, e.g.:
    a. By sex: boys, girls
    b. By seating arrangement: rows, tables
    c. By numerical groups: 1-at-a-time; 2-at-a-time; 3-at-a-time
7. All at once: Having everyone in the room go up to the problem table at one time, each person putting in his own piece
8. Same as last time: Having the group follow the plan which was followed in the preceding problem
9. Group discipline: Making suggestions about how the group ought to behave while building the puzzle, e.g.:
    a. Don't push
    b. Don't crowd around
    c. Don't jerk pieces away from other people
10. Change personnel: Suggesting new builders, new leaders, or suggesting a different order of sequential groups

The Organizational Efficiency Score provides information about how much knowledge a group has regarding ways and means of organizing itself to attack a particular problem. It will be observed that the items on the list of Organizational Techniques are, on the whole, arranged in descending order of efficiency. The first two items concern the idea of depositing leadership authority in certain group members; the next two deal with the idea of how the class should proceed if and when it is organized; the next two deal with the bases which the class can use to organize itself into groups; and the idea "all at once" represents complete lack of organization. The last three items on the list of Organizational Techniques apply only in the second and third problems. These ideas, if checked, must be evaluated in terms of the plan used in the preceding problem.

The Organizational Efficiency Score, if high, indicates that a class recognizes the need for leadership and for some means of arranging its members into a structure that will function efficiently in the present problem situation. A low score indicates that children lack the necessary knowledge of organizational procedures and can think of nothing better than to have each child put in his own piece himself.

The fourth score is termed *Task-Centeredness* and is based upon the same list of ideas as is the Organizational Efficiency Score. However, whereas Organizational Efficiency concerns *all* of the suggestions children have about ways of solving the problem, the Task-Centeredness Score is based upon only those ideas which are selected and incorporated into the final plan. The logic of this score resides in the notion that a group skilled in techniques of cooperative group action will select, from among all the ideas presented, those few ideas which are "best" in terms of the demands of the particular problem at hand. A group which is not skilled in these techniques will tend to disregard "good" ideas in favor of poorer ones.

A high Task-Centeredness Score indicates that of all the ideas a particular group of children had about ways of solving the problem the "best" ideas (e.g., those highest on the list of organizational techniques) were selected and incorporated into the final plan. A low score indicates that although a group had several good ideas they rejected them in favor of ideas which capitalized upon individual desires and minimized the total group good.

The fifth score is termed *Independence* and concerns the group's ability to plan by themselves without external assistance from or interference by the examiner. The elements of this score are based upon the kinds of actions the examiner is forced to take during the planning period and these are listed in the fourth section of your Observation Sheet.

The first action, "Prods Group," means that the individuals in a class respond with stony silence to the examiner's initial request: "Now what ideas do you have about ways the class might go about solving this problem?" When this occurs, the examiner repeats certain of the rules and the demands of the test and again asks, "Now, how are you going to get together and build this model? Does anyone have an idea of how the class might go about it?" "Prods Group" is also checked whenever the examiner has to urge the group to give *more* ideas after the first one or two.

The second action, "Suggests Vote," means that the class, after presenting a few ideas, comes to an impasse and does not know how to proceed. That is, children keep repeating ideas that have already been presented, or they again fall into a complete silence and look with bewilderment upon the examiner when he asks, "Which of these ideas are you going to select?" In either case the examiner suggests that they vote on one of the ideas which they have presented.

The third action, "Reminds Group," means that the examiner is forced to recall to the class the rules of "good group behavior." This is necessary when the children become noisy and undisciplined, when they all talk at once or shout out their ideas to the examiner. In such a case the examiner is forced to say, "Now, boys and girls, let's speak one at a time for I'm afraid I can't hear your ideas and I'm sure you can't hear each other's."

The fourth action, "Summarizes Ideas," is taken by the examiner whenever the class reaches an impasse because of the complexity of their ideas. That is, the discussion eventually comes to revolve around three or four different plans, each of which is so involved and detailed that the class cannot pull together and hold in mind all of the different elements and cannot come to any kind of decision. In such an instance the examiner names the plans in a one-two-three fashion, briefly summarizing the details of each.

The fifth action, "Conducts vote or election," means that the class has suggested that they vote on a plan or that they elect a leader. The examiner then proceeds to perform the mechanical task of conducting the vote or election.

A high Independence Score indicates that a group has the ability to discipline themselves and to carry on successfully a discussion from the beginning to a point at which decisive action is taken. Conversely, a low score indicates that a group lacks the several skills necessary for independent planning and must rely heavily on outside help to reach a point where decisive action can be taken.

The sixth and final score which is given during the planning stage is termed *Decision Method* and refers to the way in which the group decides upon a particular plan of action. These four methods are shown in the fifth section of your Observation Sheet.

The method of active consensus is that in which all members of the group come to a more or less spontaneous and mutual agreement about a plan of action. The method of active vote indicates that the class is split upon two or more suggested plans and must use the method of voting to achieve the kind of consensus necessary to group action. The method of passive vote indicates that there is no *real* concern on the part of group members about which of several plans of action is adopted. Voting is artificial and meaningless and tends to be engaged in as "something we are supposed to do." The method of passive consensus indicates that the class as a whole is uninterested in both the problem at hand and the plan which ought to be selected in resolving it.

Our hypothesis is that use of the method of active consensus indicates a warm, cooperative, cohesive and skilled group, whereas use of the method of active vote indicates a skilled but less cohesive one. Use of either of the passive methods seems to be associated with very low degrees of skill in group planning and group problem-solving.

The scores obtained during the Operations Stage have not yet been as precisely defined as those given during the Planning Stage. The reason is that the behavior which occurs during this second stage is extremely more complex and far less "controlled" than the behavior in the planning period.

The first score in the Operations Stage is termed *Execution Pattern* and indicates whether and how well the group puts its adopted plan of action into effect, once the examiner has said, "Begin.' A high score on this variable indicates that the children move easily and efficiently into the mechanics of their plan and proceed without hesitation to carry it out in detail. A low score indicates that the plan breaks down immediately and that children rush up to the problem table in a helter-skelter fashion with absolutely no regard for the rules of group behavior they decided upon during their planning session.

Our hypothesis is that children who are used to formulating the rules which are to govern their behavior and, more important, who are accustomed to *acting* in accordance with these rules will score high on this variable. Conversely, children accustomed to engaging in group planning as a relatively meaningless mental exercise will tend to exhibit some skill in the planning period but will fail completely in their ability to actually govern their behavior in accord with the rules they themselves have outlined.

The second score on the Operations Stage is as yet a highly subjective, and accordingly, a highly unreliable one. It is a measure of the overall psychological tone-quality of the group engaged in working on the puzzle. The descriptive terms listed on your Observation Sheets are checked on the average of once a minute by the observer on the basis of his judgment about what the group is like. It is our hypothesis that the "psychological tone" of the group is to some extent related to and hence indicative of the psychological quality of the classroom atmosphere which is created and maintained by the particular teacher involved. Our findings here are suggestive, but as yet so tentative that it is not possible to discuss them in detail at this time.

The final score obtained during the Operations Stage is termed *Group Interest* and provides an indication of the concern for the problem which is exhibited by those children who are not actively engaged in the construction task. We are not completely sure what is responsible for observed differences in interest in this test between various classroom groups of children, but we think that the group that maintains high interest in the task is more cohesive, more skilled in group work and more ego-involved in the test because of its being a "group project" than is the group whose interest is solely dependent upon immediate personal participation in the task.

Observer____________________

Date____________________

RSSR OBSERVATION SHEET

Teacher________________ School________________ Grade____ No. in group____ Problem____

PLANNING STAGE

1. Group Participation

Individual																	
Buzz Groups																	
Chorus																	
Noise																	

2. Communication Pattern

Suggests Discrete Idea																	
Repeats Idea																	
Evaluates Idea																	
Improves Idea																	

3. Organizational Techniques

Construction Authority																	
Supervisory Authority																	
Simultaneous Groups																	
Sequential Groups																	
Problem Groups																	
Non-problem Groups																	
All at once																	
Same as last time																	
Group Discipline																	
Change Personnel																	

4. Independence

Prods group																	
Suggests vote																	
Reminds group																	
Summarizes ideas																	
Conducts vote or election																	

5. Decision Methods

Active Concensus																	
Active Vote																	
Passive Vote																	
Passive Concensus																	

6. Description of plan:

Time Planning__________

RSSR OPERATIONS STAGE

Approach

1. Plan put into effect smoothly and without hesitation ____________
2. Plan put into effect immediately in a wild excited rush ____________
3. Plan put into effect after some hesitation ____________
4. Plan breaks down immediately ____________

The Plan

In Operation																	
Changed—Inefficient																	
Breaks rules																	

Number at Table

Plan																	
Total																	
1/2 — 3/4																	
1/3																	
Small Group																	

General Atmosphere

Quiet and Intent																	
Warm and Cooperative																	
Tense and Excited																	
Bickering and Squabbling																	
Whining and Complaining																	
Angry and Hostile																	
Instructor																	

Non-Builder Group Interest

Good																	
Fair																	
Poor																	

Remarks:

Time ____________

No. of Errors ____________

Description of Group Characteristics

JOHN K. HEMPHILL

The study of social groups and the behavior of group members is rapidly becoming an active field of endeavor for psychologists. Both empirical research and theory building are receiving attention. However, if the scientific study of social groups is to continue to be vigorous and effective, one area of problems that has been largely neglected must be given increased attention. It is imperative that a taxonomy of group characteristics be developed in a form that permits accurate measurement.

Research work on problems of group taxonomy may be less exciting than investigating the intriguing processes of group functioning, but if the many discrete research findings that are accumulating in the literature are to be brought into some order or relationship, we must use common concepts for designating major differences among groups. If we could specify the characteristics of different samples of groups by using more comparable terms than are now available, much confusion that exists at present in group research would disappear. Questions about seemingly contradictory research findings and about the generalizability of the results of laboratory research would be more susceptible to answer.

This paper reports the development and use of a tentative set of dimensions devised for use in describing group characteristics.

Development of the Group Dimensions Description Questionnaire

The following four criteria served as guides in the selection of characteristics to compose the set of group dimensions: (a) each characteristic should be meaningful within a sociological or psychological framework; (b) each characteristic should be conceived as a continuum varying from the lowest to the highest degree; (c) each characteristic should refer to a relatively molar rather than a molecular property of a group; (d) each characteristic should be relatively orthogonal or independent of all other characteristics in the descriptive system.

As a first step, and after a thorough review of relevant literature in social psychology and sociology, fourteen group dimensions were selected and defined as follows:[1]

---

[1] A more complete definition of each dimension and a detailed account of the development of this instrument is given by Hemphill and Westie. (3)

1. *Autonomy* is the degree to which a group functions independently of other groups and occupies an independent position in society.
2. *Control* is the degree to which a group regulates the behavior of individuals while they are functioning as group members.
3. *Flexibility* is the degree to which a group's activities are marked by informal procedures rather than by adherence to established procedures.
4. *Hedonic Tone* is the degree to which group membership is accompanied by a general feeling of pleasantness or agreeableness.
5. *Homogeneity* is the degree to which members of a group are similar with respect to socially relevant characteristics.
6. *Intimacy* is the degree to which members of a group are mutually acquainted with one another and are familiar with the most personal details of one another's lives.
7. *Participation* is the degree to which members of a group apply time and effort to group activities.
8. *Permeability* is the degree to which a group permits ready access to membership.
9. *Polarization* is the degree to which a group is oriented and works toward a single goal which is clear and specific to all members.
10. *Potency* is the degree to which a group has primary significance for its members.
11. *Size* is the number of members regarded as being in the group.
12. *Stability* is the degree to which a group persists over a period of time with essentially the same characteristics.
13. *Stratification* is the degree to which a group orders its members into status hierarchies.
14. *Viscidity* is the degree to which members of a group function as a unit.

As the second step in the development of the instrument for measuring group dimensions, over 1000 items, each referring to some specific characteristic of relationships among group members or to some manner of group functioning were assembled. These items were secured (a) by administration of open-ended questions about the nature of their groups to members of over 500 groups (2) and (b) by selecting excerpts from sociological, psychological, and literary writings in which groups were described.

Data for an index of "homogeneity of placement" were secured for each item as the third step.

This index was adopted to give a single numerical evaluation of each item with respect to its homogeneity, not in the usual sense of inter-item variance, but in relation to the other thirteen dimensions composing the descriptive system. We sought to find items that were clearly relevant to one of the set of dimensions and also clearly irrelevant to the remaining thirteen. Each of five judges sorted all of the items fourteen times. After carefully reading the definition of one of the fourteen group characteristics the judges placed each item into one of three categories. One category consisted of those items considered to apply directly to the characteristic. A second category included those items about which the judges were undecided. The third category was for items considered irrelevant or inapplicable to the characteristic. Upon the completion of one sort, the items were shuffled and the same procedure was repeated on the next dimension. Agreement among judges that an item applied to a dimension and agreement that it did not apply to other dimensions in the system are given approximately equal weight in the value of the index.[2]

The fourth step in the development of measures of the group characteristics consisted of a conventional internal consistency item analysis. For this purpose a preliminary questionnaire composed of the 350 items that had highest index values was administered to 200 individuals who were members of 35 different groups. These respondents indicated how well each item described their group by selecting one of five choices: Definitely True; Mostly True; Undecided; Mostly False; or Definitely False.

As a further refinement, a smaller sample of the respondents who completed the 350 item questionnaire were interviewed and questioned about why they had selected specific choices to specific items. Ambiguities in reasons for selecting responses were noted and then each item was edited to eliminate sources of ambiguity, or removed from further consideration.

---

[2] The index of "homogeneity of placement" was computed using the following formula:

$$I_i = \frac{N \sum_{j=1}^{n} X_{ij} - \sum_{i=1}^{N} \sum_{j=1}^{n} X_{ij}}{2 \left[2n (N-1) \right] + \sum_{i=1}^{N} \sum_{j=1}^{n} X_{ij} - N \sum_{j=1}^{n} X_{ij}}$$

where: j = any judge
i = any dimension in the system
n = number of judges
N = number of dimensions
X = score assigned to item placement as follows:
Definitely applies = + 1
Undecided = 0
Definitely does not apply = − 1

The final instrument entitled "Group Dimensions Descriptions" is composed of 150 statements about groups each assigned to one of thirteen dimensions.[3] A respondent's description of his group is summarized by thirteen group dimension scores. Each score is the sum of the weights given choices 'Definitely True' through 'Definitely False,' for the items assigned to each particular dimension.

Experience with the Questionnaire

The final form of the questionnaire has been used in several empirical studies since its development in 1950. Group descriptions have been secured for office and shop groups in industry (1), departments of instruction in universities and colleges (5), high school and elementary teaching staffs of 26 Ohio public school systems (8), various military units (7), religious organizations (6), sports teams, informal social groups (4), etc. We have had opportunity to examine some of the properties of the dimensions scores that are available from these studies.

Perhaps the most important question that can be asked about the group dimensions scores is whether or not different members of the same group agree about the characteristics of their group. Unless some agreement can be shown there would be cause to question seriously the existence of such an entity as a group. We have examined three sets of data where it was possible to compare the variance of the group dimensions scores for the descriptions supplied by members of the same groups with estimates of the variance of these scores based on mean scores for different groups. Significant F ratios have been found for each of the thirteen dimensions but not for all dimensions in all three studies. In a study of eight groups that were described by a total of 65 of their members, the extent of agreement among descriptions supplied by the members of the same groups, expressed as unbiased correlation ratios, ranged from .47 for the dimension Polarization to .74 for the dimension Permeability. In a second study involving 134 faculty members and their respective 20 departments, significant between-group variance was found for the dimensions Flexibility, Hedonic Tone, Intimacy, Participation, Polarization, Stability, Stratification, and Viscidity. In a third study the staffs of 26 elementary public schools were shown to differ significantly on Autonomy, Flexibility, Homogeneity, Intimacy, Stability, and Viscidity. It appears evident that the descriptions supplied by members of their groups with the use of the Group

---

[3] Size was dropped from the final form of the questionnaire, not because it was considered an irrelevant dimension but because there appeared to be more direct methods of obtaining this information.

Dimensions Description Questionnaire provide differentiations among groups.

Some evidence of the validity of such differentiations has also been observed. For example, descriptions of a military training department (ROTC) were contrasted with descriptions of a committee from a college of education and the expected large differences along the dimensions of Control, Participation, and Stratification clearly emerged.

Not all of the variance among respondents who describe the same group is to be attributed to errors of measurement. The dimension scores have been shown to be related to characteristics of the respondents' status in the group. For example, in the study of the departments of a liberal arts college it was discovered that those members of the faculty who possessed greater status, i.e. held the rank of Assistant Professor or better, described their departments with lower scores on the dimension Control than did other members of their department, who were largely Instructors. In the same study it was also found that those individual members of the faculty who expressed above average satisfaction with their jobs tended to view their departments as higher on the dimensions Hedonic Tone, Participation, Polarization, Potency, and Viscidity, but as lower on the dimension Stratification. In a study involving aircrews, which used a modified form of this questionnaire, Rush (7) found that Aircraft Commanders differ from other crew members in the way they describe their crews. No study has been made relating measures of the personality variables of respondents to their descriptions of their group, but such relations might be expected. Thus we see that the group dimensions scores are sensitive both to commonly seen group characteristics and to the respondent's particular relationship to his group.

Considerable effort in the development of the Group Dimensions Descriptions questionnaire was expended in attempting to secure independent dimension scores. In order to examine the extent of overlap among the dimension scores empirically we have computed intercorrelations for three sets of data; a sample of 100 descriptions of miscellaneous groups, a sample of 136 descriptions of 22 departments of a liberal arts college, and a sample of 315 descriptions of the staffs of 26 Ohio public school systems. In general the correlations among the dimensions are quite small—approximately three quarters of them below .3 in magnitude. Certain dimensions are frequently found to be correlated more highly with each other. For example, the relationship between Hedonic Tone and Viscidity was found to be .33; .64 and .57 for the three samples respectively; between Participation and Polarization, .09; .34 and .47; and between Stratification and Autonomy, − .54;

− .31 and − .24. The size of the correlations between specific dimensions appears to be sensitive to the institutional setting from which the groups are selected. Experience accumulated to date has provided no clear-cut evidence for the abandonment of any one of the dimensions if they are to be considered applicable to all varieties of groups. However, it does appear that if the investigator is interested only in differences among a relatively homogeneous sample of groups, certain of the dimensions might not be pertinent. For example, in adapting the instrument for use with aircrews, we eliminated the dimensions of Size, Autonomy, and Permeability, because the assumption of variance on these dimensions appeared far-fetched.

Finally, the problem of establishing normative data for these group dimensions has been given attention. I frankly do not know how to establish defensible norms for group characteristics. The problem appears to be one of lifting yourself by your bootstraps. Certainly, securing representative samples of individuals would not insure a representative sampling of their groups, since individuals have numerous and complicated multiple group membership.

Conclusions

The development of a tentative set of dimensions of group characteristics has been described. Some of the experience that we have had with the use of the dimensions has been cited briefly. Much remains to be done. We need defensible norms. We need to reconsider the selection of group characteristics. Perhaps the list of dimensions should be enlarged, or perhaps it can be reduced. But unless we solve some of these elementary problems of describing the basic entity of research on groups, confusion will reign.

REFERENCES

1. Gekoski, N. The relationship of group characteristics to productivity. Unpublished doctor's dissertation, Ohio State Univer., 1951.
2. Hemphill, J. K. Situational factors in leadership. Bureau of Educ. Res. Monog. No. 31, Columbus: Ohio State Univer., 1949.
3. Hemphill, J. K. & Westie, C. M. The measurement of group dimensions. *J. Psychol.*, 1950, **29**, 325-342.
4. Hemphill, J. K., Seigel, Ann, & Westie, C. M. An exploratory study of relations between perceptions of leader behavior, group characteristics, and the expectations concerning the behavior of ideal leaders. Unpublished manuscript.
5. Hemphill, J. K. & Bentz, J. Leadership behavior associated with the administrative reputations of the departments of a college. Unpublished manuscript.
6. Knight, R. A study of thirteen group characteristics of selected religious organizations at Ohio State University. Unpublished master's thesis, Ohio State Univer., 1950.
7. Rush, C. Group dimensions of aircrews. Unpublished doctor's dissertation, Ohio State Univer., 1953.
8. Seeman, M. A sociological approach to leadership: The case of the school executive. Unpublished monograph.

DISCUSSION

PARTICIPANTS

William E. Coffman, Richard S. Crutchfield, Dora E. Damrin, Hugh M. Davison, Lindsey R. Harmon, John K. Hemphill, Irving Lorge, Elliott M. McGinnies, Emma Spaney, S. David Winans

Dr. Davison: I have a question for Dr. Hemphill. Say you have a group, and it has dimensions on your test over here to the right on a distribution. You have another one over here to the left on a distribution. Now, how do I know whether or not, if I traded people around, they would not coincide? Is it possible they are all making their own frame of behavior and therefore, might it be impossible to really compare groups?

Dr. Hemphill: You are asking a question which I do not know that I can answer. I would suggest one thing, however. You might approach such a problem by using the questionnaire that I have described. Off hand, I would not attempt to answer the question of whether we change the group when we change some or all of the persons in it. I would have to know much more about the group before I could even make a guess. I think you need a tool or instrument like the Group Discussion Questionnaire.

Dr. Lorge: I was hoping that the other question would be asked: if the individual could identify the groups to which he belonged, could you then specify the groupness of the group of which he is a member, and if there be homogeneity in his group memberships?

Dr. Hemphill: I think I would have to give the same answer to your question, Professor Lorge. We could study the problem. I do not know what the answer might turn out to be. I should like to re-emphasize, however, that we are quite convinced that we are studying something involving groupness when we are able to show that the members of a group give descriptions of their group which differ from descriptions of other groups by their members.

Dr. Winans: I have two rather simple questions. First of all, I would like Dr. Damrin to show us the other problems she mentioned; and second, I would appreciate it if Professor Hemphill would comment on the difference between his items *permeability* and *stability*.

Dr. Damrin: As I told you, the problems are designed so that they increase logically in order of difficulty. In the first problem, the house, you will notice that the colors are all together in straight lines and that the square and triangular blocks are not inter-mixed.

In the second problem, the bridge, the colors are still separated, but the triangles and squares are mixed, and are thus more difficult to put together.

In the third problem, the dog, both the colors and shapes of the blocks are inter-mixed, thus making it the most difficult of all. On this problem, by the way, the children are asked to beat their time on the bridge problem. That is, they do not get fifteen minutes to do it, they get only the number of minutes it took them to build the bridge. This forces them to evaluate and improve their procedure.

Dr. Hemphill: I presented very short definitions of these dimensions. I think a little elaboration would make clear the difference between stability and permeability. Permeability refers to a characteristic of the group which is often reflected in membership requirements, or the difficulty of getting into the group—what you have to do or be in order to qualify for membership. Stability refers to whether the group remains the same with respect to all significant characteristics over a period of time—in other words, a stable group in 1950 might be exactly like it was in 1899.

Dr. Spaney: I should like to ask Dr. Damrin whether she plans to extend her experiment to include verbal, or does she plan to stay with what she has. I am thinking particularly of adult problems.

Dr. Damrin: At the present time, we do not plan to extend the construction type of task and develop a set of tests that are more verbal. The chief reason is that verbal tests depend rather heavily on intelligence, and we are trying to keep intelligence out of the picture as much as possible. These problems are sufficiently simple that even classes of very dull children can do them. In fact, we have successfully administered them to classes with an average I.Q. as low as 90.

In the adult area we have found that the test as it now exists is quite appropriate, although the present scoring system is not completely so. I have given the test to my research board, all of whom are Ph.D.'s, and in general the obtained results were quite satisfactory in terms of our present scoring system. It is our feeling, therefore, that to go into the area of verbal problem solving would serve primarily to add unnecessary complexity, and the test, as it now stands, already has so much complexity that it is extremely difficult to score.

Dr. Harmon: I should like to ask Dr. Crutchfield if he has any information on the validity of these conformity scores against criteria of importance to the individual's adjustment on the job, or anything of that kind, other than ratings or other test scores.

Dr. Crutchfield: We do have some external criteria of performance in the study of professional men. Here, the zero order correlations with conformity scores do not seem very promising. There are, however, questions of whether the criteria are adequate and how they should be predicted to relate to conformity. All that I can really say is that we have some work in progress on this point, but no clear results as yet.

Dr. Coffman: I should like to ask Dr. Damrin a question. She has made the statement in the general presentation that she has no data on validity or on reliability and did not add the statement, "of the traditional sort." There are three questions involved here: one is, does the behavior of the classes on this test relate to any other evidence of what the group problem-solving ability of the class is? Second, does the test differentiate among classes? And third, can scorers see the same thing? Do you have any evidence on those questions?

Dr. Damrin: My statement that I have no data of a formal type referred to the fact that I have no tables of correlation coefficients to show you.

In answer to the first question, the pretrial study on this test was done in Trenton, New Jersey. The elementary school supervisor there knew her teachers very well and selected for us, but unknown to us, pairs of teachers. That is, from any particular school she would pick one teacher who she thought was trying to get across in her classroom the things which the Russell Sage test is concerned with, and for the other she would select a more formal type teacher. In every instance but three, the tests discriminated these two types of teachers. In other words, in the more formal class there would tend to be very low group participation; in the other, greater participation. In one class, student leaders would tend to emerge; in the more formal class this would not occur, and so on. Therefore, we have those kinds of evidences about the validity of the tests.

To the question of does the test discriminate between classes, the answer again is yes. Preliminary scaling of the scores that I have described to you reveals that classes differ very markedly from one to another, and yet the same class is relatively consistent in its behavior throughout the three problems. This gives us one type of evidence about test reliability.

To the third question, do scorers see the same thing when observing children on the test, I can only answer that I hope they do. I have just finished training two teams of examiners in New York. For the planning stage we seem to be getting reliability that I would roughly estimate is about .9. This is on the basis of one week's training. On the operations stage I am afraid the reliability is about .5. This is because factors such as the "psychological tone quality" of a group of children require much more actual experience to evaluate accurately than do factors such as "amount of group participation."

My present feeling is that the data for the operations stage are not yet usable because of this fact.

Dr. McGinnies: Dr. Crutchfield, I was interested in your description of some of the parameters that seemed to influence conformity behavior, particularly those measures of personality characteristics, and it sounded as though the nonconformists had all the desirable personalities while the others had the undesirability traits. If this is true, I wonder whether you could say a little bit more about some of the instruments that you used to arrive at these conclusions.

Dr. Crutchfield: I should emphasize that the particular way the studies have been conducted has not been such as to measure extremes of nonconformity, that is, negative suggestibility effects, so to speak. This could easily be done. There is some incidental evidence in the results which would lead one to suppose that overinsistence on independence may at times be associated with undesirable personality features. But this has not yet been fully investigated.

Coming to the main point of your question, the instruments used in the personality assessment are very varied. The assessments of the professional men and of the mature women lasted three full days, and involved such procedures as standard tests, specially designed personality inventories, group interaction procedures of various kinds, interviews, projective techniques, physiological and medical measures, and so on; in short, the usual range of things found in large-scale assessment. The correlations I was reporting are drawn from all of these.

One representative finding is from the Terman Concept Mastery Test, a newly devised instrument measuring superior intellectual performance, which correlates approximately .50 with independence.

Some of the scales of the MMPI and similar personality inventories were also used. Another very important source of information came from Q-sort descriptions of the assesses made by the assessment staff in complete ignorance of their performance in the conformity procedure. Here there were found quite striking relationships along the lines I outlined.

Finally, since these days the famous or infamous F-scale studies of personality are much in the literature, I should report an appropriate finding there, that correlation with conformity is of the order of .40.

Dr. Lorge: I take the Chairman's prerogative in summary. The three papers, by design, make a beautiful unity in this important area. They give a sagacious attempt to get at a concept of the nature of the group, with the possibility that it will enable workers to interpret the degree to which published evidence about groups really represents integrated and functioning groups. Perhaps I am too disturbed by *ad hoc* groups created in classes by instructors saying, "The next five individuals will constitute group 1 and the next five group 2, etc." for purposes of measuring group performance or group productivity. They show that in some process observations it is quite clear that group behavior is related to the fact that some members may not work at all.

And the papers show interesting dual relationships in attempts to evaluate group process as well as the group product. The complexity of relationships exhibited are: (1) between kinds of groups, (2) between kinds of members, (3) the nature of the processes related to groups and members, and (4) the quality of the product, whether it be a solution, an action program, etc.

A large number of theoretical issues come to the fore, because as far as I can tell, the evidence about the relationship between group process and group product is just emerging. More observations and data are needed, not only in the area of action, but also in the area of the quality of the group solutions as opposed to individual solutions. The distinction betwen the productivity of individuals and of groups must be made.

ADDRESS

.... And Have Not Wisdom

DANIEL STARCH

When your chairman invited me to address you here today, he suggested that I need not necessarily talk about technical problems but that I might discuss anything I thought relevant to the purposes of your conference. I shall, therefore, deal with some problems which have been forced upon us by the events of the last fifty years. They are especially vital to all of us engaged in the study of human behavior. The solution of these problems will most assuredly affect our well-being. Failure to solve these problems aright is certain to affect volcanically the culture and welfare not only of ourselves but of many generations to come.

The problems to which I refer have become particularly acute in our time because they arise out of the wide, yawning gap that exists today between the tremendous advance in science, particularly applied science, and the appalling backwardness still manifest in human relations. Our advance in scientific knowledge has been so rapid that the dreams of Jules Verne have seemingly come true, but our backwardness in humane behavior, in dealings between individuals, between groups, between peoples and nations, has been so retarded that it is open to question whether there has been any progress at all in the general spiritual level of human relations since the Greeks took Troy with their wooden horse.

First then, let us look for a moment at the growth of science. Interest in scientific method began to germinate, to be sure, as far back as Aristotle, but more especially three centuries ago when Francis Bacon wrote his *Advancement of Learning* (1605). But it is only during the past fifty years that the great burst in the growth of science has occurred. I think it is fair to say that more advance in scientific knowledge has occurred during the last fifty years than in all previous centuries combined. We are all aware of this in a general way, particularly those of us whose personal recollections go back to the turn of the century. We know from our own observation the fantastic changes in the number of telephones, automobiles, machines, and gadgets of all kinds. But let me deal briefly, though in reasonably specific detail, with a few fields that may serve to dramatize this unprecedented grow[illegible]

Take, for example, travel and transportation. When George Washington was President of the United States, it took longer to travel from his home at Mount Vernon on the Potomac to New York than it takes today to fly from New York to Bombay. If we visualize the earth in 1800 as a globe eight feet in diameter, this globe in terms of travel time had shrunk by 1900 to the size of a basketball and by 1954 to the size of a golf ball.

This phenomenal advance in scientific know-how is fundamentally associated with the development and use of energy. Man has been able to increase his productivity, not by increasing the strength of his arm and back, but by the use and control of energy in the form of tools and machines. Around 1800 steam engines were just coming into use. By 1850 the United States had two million horse power, or one-tenth of one horse power per capita. By 1900, it had risen to one-half a horse power per capita and as of today the total energy in all prime mover engines in the United States is over six billion horse power, or about 39 horse power per capita. (Estimated by John A. Waring, Jr., *Steelways Magazine*, August 1954.) This is eighty times as much mechanical power as was available only fifty years ago. Productivity per man hour in manufacturing industries has increased fivefold since 1900.

This enormous advance in technology has been stimulated by, and made possible by, the manifold increase in technical education. Around 1900 there were 1,150 electrical engineers. Today there are 46,000, a forty fold increase. In all fields of engineering combined, there were at the turn of the century about 13,000 members in the various professional engineering societies. Today there are 170,000, or thirteen times as many.

But this is not all of the broad picture of rapid change. Take the study of man himself, his characteristics, behavior, and motivations. Here also in the last fifty years, a tremendous acceleration in scientific research and knowledge concerning human behavior and human relations has taken place. To cite merely one example, around 1900 there were only about 125 members in the American Psychological Association. Today there are more than 12,000, a growth in technically trained personnel exceeding even the relative gain in members of professional engineering societies. Add to that similar growth in number of trained economists, sociologists, political scientists, and historians and you have a truly impressive army of professional talent devoted to human activities and relations, whereas fifty years ago the picture was largely one of solitary pioneers.

Take another aspect, channels of communication, messages and materials for influencing people's behavior have increased at an unprecedented rate both in numbers and in speed of transmission. Today there are in the United States one and one-fourth copies of newspapers per day per family, two and one-half copies of magazines per week per family, two radios per family, and one television receiver per two families.

Mass communication has so speeded up that it is almost instantaneous. An important event occurring in Washington at breakfast time will be known in all important capitals of the world before your toast has cooled off. By contrast, when Lewis and Clark arrived at St. Louis in December 1803 to start their westward exploration, the Spanish commandant stationed at St. Louis refused to let them cross the Mississippi because he had not yet received official word that the territory west of the Mississippi had been ceded to France and sold to the United States months before.

Look at still another area—education. School enrollment has taken an immense upward surge. The number of pupils and the number of years in school have increased by leaps and bounds in the last fifty years. In 1900, the total enrollment in our high schools was 700,000. Today it is 7,000,000. Around 1900 the number enrolled in our colleges and universities was 115,000. Today it is 2,500,000. In other words, high school enrollment has increased tenfold and college enrollment more than twentyfold. Put still another way, 80% of our young people of high school age are enrolled in our high schools and 30% of college age are enrolled in our colleges. Hence, as of now, of all adults 25 years or older, 60% have attended high school of whom one-fourth have also attended college.

Now let us take a searching look at what has happened in the management of human affairs. Has there been any improvement at all in the spiritual level of society in the fifty years during which these marvelous advances in the physical and cultural standards have occurred, or even during the three centuries since the days of Francis Bacon, or if you will, since the days of Aristotle, Moses and Confucius? Here we are in a field that is without a generally accepted measure of value. But speaking in all humility, I for one doubt that we have made any real progress. I doubt whether we have learned to live more wisely. We certainly have not learned to live in peace. On the contrary, wars have not diminished in frequency and have become more brutal and destructive than ever. The first World War cost the United States alone 30 billion dollars and the second World War 380 billion dollars. But far worse, of course, is the destruction of human life. During the Thirty Years War three centuries ago, one-third of the population of Europe died. During the first and second World Wars lasting nine years, twenty-five million people were killed. It may well be that more people have perished through war in the last fifty years than in all previous known history, and these are the fifty years during which has occurred our boasted rise in technology and know-how, both in the physical sciences as well as in the human behavior sciences—in schooling and in the means of mass communication for influencing men's minds. Can it be that we have not realized that with increase in technical knowledge there must be a corresponding increase in responsible humane behavior if wholesale brutalization, nay even universal destruction, is to be avoided?

Turn to the behavior of persons as individuals in their family and community relations. According to United States government statistics, there was, in 1900, one divorce to thirteen marriages. Today there is one divorce to four marriages, a ratio three times as high today as fifty years ago. According to F. B. I. records, offenses including murder, manslaughter, rape, robbery, assault, burglary, larceny, and auto theft increased 37% in the thirteen years from 1939 to 1952.

Why is there so wide a gap between what we know and what we do with what we know, between knowledge and the wisdom of how to use that knowledge? There are at least three reasons. The first is that many people do not recognize that knowledge and use are two different things. Knowledge as such does not guarantee right use, in fact it does not necessarily assure use of any kind. Only purpose can determine use. The invention of a method of making fire was a landmark in the march of man, but knowing how to make fire did not determine the purpose for which the fire was to be used. It could be used to warm the hut and cook the food, or it could be used to burn down the neighbor's hut and destroy his crops. Whether nuclear energy will be used for man's benefit or for his destruction will depend on purpose. Science provides technology but not teleology.

The second reason is a by-product of the first. Technological advance makes comfortable, luxurious living an easy choice. This tends to push aside long range consideration of wise living. We feel so comfortable, so luxurious, so smart that we falsely feel self-sufficient, superior and trouble free. As necessity is the mother of invention, so trouble is sometimes the mother of humility.

The third reason is that knowledge generates power, and power over the forces of nature, and particularly over the behavior of man, whets the appetite for more power, for going rough shod over the sensibilities and rights of others. Unless power is accompanied by an increase of sensitivity to the common weal, it ends in abuse, oppression and brutality. As Lord Acton put it, "Power corrupts, absolute power corrupts absolutely."

The net effects of our fabulous advance in science are three things: we work less, about half as many hours as formerly; we live longer, about 70 years against an average of 25 years; and we live more comfortably. *Quaere,* however, whether we live more wisely and more happily. We worship the god of science in the laboratory but unfortunately neglect the goddess of wisdom in the temple. As an interesting side-light, consider the case of the five employees, all foremen, sent in 1953 by the N. V. Nederlandsche Linoleumfabriek of Amsterdam to work at regular American wages three months in the Armstrong Cork Company plants in Pennsylvania. They came not to learn new skills or to obtain technical knowledge of linoleum production but solely to get the "feel" of America. These men were greatly impressed by the facilities, the ownership of automobiles, and particularly the amount that they were able to earn and save during their short working period One man commented that he saved enough to buy a good used automobile and would have bought one if he could have driven it all the way back to Amsterdam. He questioned, however, whether an American is any happier with his automobile than a Dutchman with his bicycle.

What remedy is there? How can advances in science be made to produce advances in wisdom? The remedy is not, I am sure, as someone suggested, to halt research in order to give ourselves a chance to catch up with our knowledge. That would merely halt the progress of science without any gain whatever in wiser living. H. A. Overstreet has remarked, "Better a dish of illusion and a hearty appetite for life than a feast of reality and indigestion therewith." But this alternative so vigorously expressed is by no means a necessary alternative. By all means let us continue research and more of it, but let us also search to find out what needs to be done to live more wisely. Therefore, I wish to suggest for your consideration a three point plan of action to close the gap between the advance of science and our discouraging delinquency in human relations.

Having recognized that knowledge of itself does not automatically lead to wise use, let us then first of all *make it a specific goal to seek wise use of scientific knowledge.* Psychological research has shown that training in precise, logical thinking with mathematical concepts does not automatically lead to more accurate, logical thinking in other areas and with other data and concepts.

It is essential, therefore, that all people of goodwill keep their goal of constructive use of science bright and shining both for themselves and for others. Those with evil intent keep their minds firmly fixed on their goals. But people of goodwill are more likely to go along passively without much thought about their direction. The goal-conscious persons, whether their goal be good or evil, have this dynamic advantage: They know where they want to go and they keep going.

What *is* wise use of knowledge and how can we determine whether a given use is wise? The decent, self-respecting life, whatever else it may require, has this one absolutely unfailing requirement: *Freedom of choice,* freedom of choice in thought, attitude, speech, action and occupation, and with it freedom to seek satisfying reward according to effort, ability and achievement. In order to give proper regard to the rights and opportunities of others, in order that everyone may have maximum freedom of choice in our complex world, let us call it the *optimum freedom of choice.*

The second point of action that I submit for your consideration is that *we must learn and teach how to make sound value judgments.* Discoveries of science and uses of knowledge can and must be evaluated by their contribution to the optimum life, toward life at its best with its maximum freedom of choice. Specifically, how can we do this? Let me suggest three criteria.

1. The *test of cause and effect.* Human behavior events are linked to each other by cause and effect. They do not just happen from nowhere. They arise out of previous events and actions combined with human motivation. The first test, then, that I wish to propose for sound evaluation of ideas and actions is to ask this question, What will be the effect of the idea and of the proposed course of action based on it—not merely the immediate effect but especially the long range effect? One of the distinguishing characteristics of man is to be able to think ahead, to project himself into the future. But so much of our thinking is entirely in terms of immediate effects when in fact the important dynamic effects are the delayed, long range consequences. The immediate effects may be "sweet as honey in thy mouth" but the delayed effects may be "bitter in thy belly."

Two things are essential in visualizing the chain of cause and effect: full information and objective interpretation. These are particularly urgent when dealing with current problems and events as they are reported through the many channels with all the screening and slanting along the way. As an editor said, he did not care whether people read his editorials on the masthead page, he put his editorials into the headlines on the front page. To test the extent to which news is slanted and screened, compare the same events as they are reported in different newspapers, magazines and over the air. Note especially the differences in the headlines and lead paragraphs. You will see what is omitted, what is emphasized, and what is slanted. In a free society we cannot and must not restrict news. Read and listen to all sides and sources and then ask, What seems nearest the truth and the right?

2. The second criterion for making sound value judgments is the *test of spiritual survival value*. Will the proposed course of action in the long run help or hinder, build or destroy, the optimum life? That cannot be done perfectly since it means looking into the future, but certainly it can be done far better if we set our minds to it.

3. The third and crucial criterion is the *test of personal substitution*. Are you willing to put yourself in the place of anyone else affected by the proposed course of action? Theorists and leaders may very well think they know what is good for everybody else so long as it is the other fellow who is affected, but no course of action is good unless it is good for both leader and led. In a dictator state, would the driver be willing to take the place of the driven? When the court in a recent trial gave the persons convicted of communism the choice between going to jail or going to Russia, they chose going to jail. This third test is obviously the golden rule, with the phrases reversed. Is the leader willing to have done to him what he proposes to do to the led. Human experience has found no better test.

I make no apology for stressing these points. The events of our times have forced them upon us. It is the especial province of psychologists, of all students and researchers in the human behavior sciences, economists, sociologists, political scientists, historians, journalists, but especially psychologists, to assume responsibility for closing the gap between right and wrong use of scientific knowledge. In his recent annual report, Dr. James R. Killian. president of Massachusetts Institute of Technology, urges a course half engineering and half humanities and social sciences and stresses the clear need for a larger place in the curriculum for psychology. But promoting research and providing instruction in these fields alone will not achieve the result. The key lies in the spiritual survival values, in understanding, appreciating and cultivating them.

And this brings me to the third point in my program of action: *Train young people in making value judgments as part of our educational system*—not necessarily as separate courses, although one course might well be established, but as parts and aspects in all courses dealing with human problems—and what courses do not? It may well be not only a part of courses in the human behavior areas—psychology, economics, literature, sociology, history, and mass communication, but also in the physical, chemical and biological sciences. Training in making value judgments may follow procedures such as I have just outlined. Right value judgments and attitudes will not be developed unless conscious effort is made, and made with the materials in all areas of study and research. As psychologists, you know that we learn little from passive exposure.

Our day-to-day behavior is a mixture of (A) voluntary, purposely directed thoughts, attitudes, and actions, (B) involuntary, externally, and forcibly directed thoughts, attitudes, and actions, and (C) conditioned, habituated thoughts, attitudes, and actions. When "A" and "B" are repeated often enough, they drop down to the "C" level and automatically condition our lives. "A" is the avenue used in a free society. "B" is the whip used in a dictatorial society. It is at the "A" level that we as free individuals have maximum control of our own lives.

If the present rate of school enrollment continues for another generation, eight out of ten adults will have gone to high school and three out of ten will have gone to college. With this upsurge in school attendance, it is all the more urgent that education in long range spiritual survival values become a part of the cultural climate of our time. It is shocking to discover the utter illiteracy and immaturity in the thinking concerning the durable human values on the part of so-called educated, intelligent people.

Here is one example. Why should anyone get more than $7500 income a year? Analyze this through. Apply the criteria I listed a moment ago.

1. What would be the effect on society as a whole? On individual effort and incentive?
2. Would these effects help or hinder life at its best with optimum freedom of choice for everybody, not merely for the individuals concerned?
3. Would the enforcer of the idea be willing to exchange places with anyone else affected by the plan?

To think it through thoroughly and objectively would be good training in value judgments not only for teenagers but for many adults.

Take another idea—not prevalent today but held stubbornly a hundred or more years ago. Owners and managers of textile mills made it so difficult for adults and children to hold their jobs in the mills that it was next to impossible to get time off for children to go to school.

Again apply the same three tests. The stubborn obstinacy on the part of the mill owners led to the formation of labor unions—and that was a good thing. It was possible thereby to throw off this oppressive yoke. In turn, however, power of the union leaders led to extremes on the other side, as always happens. Excessive power at one extreme generates counteracting forces which in turn lead to excessive power and abuse on the other extreme.

As I look back upon my own student and teaching days in several large and highly respected institutions, I recall how professors would shy away from value judgments, from moral and spiritual considerations as something on a level with old wives tales.

Science has in its hands the power of life and death. This is literally true not only in the physical area but in the psychological area as well. Look at the subtle mind conditioning and the brutal brainwashing and the frustration of the free spirit of man that we have seen going on in the last 25 years. Wars today are not won on the battlefield. They are won in men's minds before the fighting in the field even starts. I propose, therefore, to you psychologists especially interested in educational testing problems two things:

1. That you urge the need for training in value judgments as part of our educational processes, and

2. That you devise tests for measuring value judgments and attitudes and changes produced by such training.

The far-reaching effects of many ill-considered, not to say irresponsible, decisions of men who govern so greatly outweigh the passive goodwill of the millions who suffer therefrom that the only cure can be in the goodwill of the millions to become directly active in understanding and cultivating value judgments and attitudes. A dictator can do with the stroke of a pen what millions know to be wrong but are helpless to do anything about. As Dr. Albert Schweitzer recently said, "Scientists must speak up."

And now to summarize briefly:

1. Recognize that knowledge of itself does not automatically lead to wise constructive use and that therefore we must set about specifically to seek and promote wise use.

2. Determine what is wise use by evaluating proposed ideas and courses of action against three criteria such as I have outlined:
   a. The test of cause and effect.
   b. The test of spiritual survival value.
   c. The test of personal substitution between leader and led.

3. Make the training of young people in value judgments a consciously recognized part of our school and college teaching.

It is evident what will decide our destiny, our ways of living, our happiness, our well-being or our misery, will not be the miracles of science by themselves but the stamina of our wisdom in using this knowledge. "Wisdom and knowledge shall be the stability of thy times, and strength of salvation."

GENERAL MEETING

New Developments in the Education of Abler Students

Acceleration: Basic Principles and Recent Research

SIDNEY L. PRESSEY

Sometimes a topic becomes so involved with situations emotionally charged that dispassionate consideration of it seems almost impossible. Occasionally an unfortunate label leads thinking astray. Educational acceleration is a subject which has suffered greatly from these handicaps—and more. It was closely associated with the war and two clumsy expedients then: college entrance without completion of secondary school, which aroused bitter antagonism in public school people; and lengthened college year, which burdened and antagonized college faculties. "Acceleration" implies hurry with probable superficiality. And every experienced educator has known bright youngsters double-promoted into an older group who there felt miserably out of place—perhaps he had been such a case.

In contrast, this paper will beg your open-minded consideration of the possibility that for abler students to progress through school at faster-than-average pace is normal *for them,* not hurrying; that there are ways of facilitating their progress which help rather than hinder good social adjustment; and that such steps can lessen the load and facilitate the work of our overcrowded schools. An additional gain should be very timely. Russian universities and technical schools appear now to be graduating about three times as many engineers and twice as many scientists a year as American institutions.[1] Multiple evidence indicates that facilitating the progress of able students leads more of them to complete collegiate and professional training; also probably the occasional notable genius (the Edisons or Einsteins) will thus be more likely to reach full fruition.

Two Major Neglected Facts of Human Development

Of basic importance are two conclusions from recent developmental studies regarding the gifted—in childhood and youth, *and* in adult career. First, gifted children tend to develop more rapidly than the average youngster, not only intellectually and educationally but also physically and in personality. The bright six year old is likely to be not only in intelligence, but also in reading and physique and social assurance, more like a second than a first grade child. The bright sixteen year old is probably not only in ability and in general knowledge up with the general run of eighteen-year-old high school seniors or even nineteen-year-old college freshmen; he probably reached puberty earlier than average and is in physique, interests, and social adjustment more mature than the average for his age. To have him in the eleventh or twelfth rather than the tenth grade, or to start the bright six year old in the second grade, is not to hurry him but rather to have him progress according to his real growth rate. Terman remarks that, "The exceptionally bright student who is kept with his age group finds little to challenge his intelligence and too often develops habits of laziness that later wreck his college career. I could give you some choice examples of this in my gifted group."[2] There is also evidence that holding a bright youngster back with his age group is less favorable to good social adjustment than carefully advancing him into a group more like him in ability and maturity of personality.

---

[1] Meyerhoff, H. A., United States Shortage: Scientists, *U. S. News and World Report,* January 15, 1954, 46–9.

A second major question of educational policy regarding the able student is usually not faced squarely: if he is to be most productive in career and most adequate as a citizen, at about what age should he be through with full-time school and really begin his adult life? Vital statistics show the late teens and early twenties to be the healthiest years; physical tests and athletic records show them to be physically the most vigorous. And Lehman's very extensive findings regarding ages of most brilliant scientific discoveries, most important inventions, best writing, most remarkable paintings, all indicate that the most outstanding creative work is done early in the adult years—often in the twenties, sometimes even in the teens.[3] The total of such evidence is impressive. Nevertheless, American educational programs for our most able young people are being more and more lengthened; they must have not only college but also graduate or professional training, perhaps an "intern" experience—possibly even some post-doctoral work. Before the second world war, the median age of receiving the Ph.D. degree in this country was 30; now, it tends to be a little later. By the age of 25, Edison and Einstein were doing important creative work. If they were of that age in this country now, they would instead probably be worrying about their language requirements! However,

---

[2] Terman, L. M. The discovery and encouragement of exceptional talent, *Amer. Psychologist*, 1954, 9, 221–30; also *The gifted child grows up*. Stanford Univer. Press, 1947.

[3] Lehman, H. C. *Age and achievement*. Princeton Univer. Press, 1953.

Edison might have been spared such difficulties; he would probably not have been admitted to an American university, not having even been to high school!

The argument thus is that *able students should progress more rapidly than the lockstep rate through school and college because they develop more rapidly than the average young person, and should get into their productive careers earlier than occurs with the lockstep.* But what about possible social maladjustment, gaps in training, or damage to health?

Means of "Acceleration"

By a historical perversity, the worst means for rapid progress—grade-skipping in school and the lengthened school year in college—have been so much more used than better methods that, to many people, these worst ways are synonymous with "acceleration." The boy who is skipped from fourth grade to sixth (obviously a half-grade skip is less risky) may suffer at least briefly from ignorance of some arithmetic process taken up in the fifth grade. If he is a conspicuous only one thus to be advanced into a close-knit little sixth grade social unit, he may initially meet hostility—and parents may talk. But even this clumsy method — grade-skipping — need not cause much trouble. Bright children are usually ahead of their age in their reading and other subjects; a little help from a teacher or parent usually takes care of any omissions. If grade-skipping of bright pupils is made fairly common and if (as is common nowadays) changes in the membership of a grade group are frequent for other reasons, social difficulties are usually minimal. Trouble is likely only if the "accelerated" child is not really of superior ability but is pushed ahead because of parental pressure, or school work neurotically superior in compensation for a social maladjustment already existing. Grade-skipping is the easiest method of advancing a pupil, especially in a small school; and if "skippers" are selected carefully on the basis of adequate measurement (and a little help is given in adjusting for the skipped work and to the new group), outcomes are usually good—as will be seen shortly.

However, there need be no skipping over possibly needed school content, or moving an occasional child into a strange group. Children may be admitted to the first grade (or transferred from kindergarten) on the basis of mental age and reading readiness rather than chronological age; and some half-dozen investigations have all shown bright five year olds, so admitted, thereafter excellent in school work and in relations with other children. Or bright six year olds with some initial skill in reading may be started in the second grade. Or, a "primary

105

pool" may throw together all children usually in the first three grades, and move each on into the fourth grade when he is ready. Large junior high schools may have rapid progress sections, made up of bright youngsters with excellent school records to date and good social adjustment, which do the usual three years' work in two. And several careful investigations have shown these rapid-progress pupils doing as well in senior high school, academically *and* in relations with other students, as pupils of the same general ability and total record at the beginning of junior high but who spent the usual three years there. Similar rapid progress sections (three years in two) in senior high have shown similar success in college.

A few college students may be really fatigued by a lengthened school year, though a total of six weeks or more of vacation usually still remains; both fatigue and protest seem greater in the faculty. A few students who need the earnings or the experience of summer employment may miss these benefits. But the questionableness of a double assumption in the method must be stressed: that bright students can learn only in college courses, and that they must spend as much time in these courses as average students. A substantial number of studies have shown that students who obtained credit for a course by passing a comprehensive examination in it instead of going to class (preparation having been obtained from superior or extra work in secondary school, from independent study, travel, or otherwise), do excellently in later courses in that subject and also in total college record. Streamlined sections of college courses, with methods adapted to the superior student and with reduced hours in class to facilitate acceleration, have been reported as successful, also honors programs combining or replacing several courses, and guided independent study. *In short, the progress of superior youngsters may be facilitated in a variety of ways; most of them are better than grade-skipping or lengthened school year.* It need hardly be mentioned that wise use of each requires initial testing to assure superiority, and continuing measurement thereafter to guide progress and determine outcome.

Whatever the methods for rapid progress, modifications should be attempted of the present powerful social pressures for the lockstep. Thus, from the day of entrance, the college student is designated a member of the class which is to graduate *four* years later; the entrant in '54 is a member of the class of '58 (English universities are wiser—there, an entrant in '54 is a member of the class of '54). Each of the four years is given a distinctive name and status; special opportunities (as for presidencies of student organizations) are open only to students in their fourth year. A student who finishes college in three years soon becomes a social anomaly who belongs nowhere. Instead the feeling should be fostered of belonging to the school not the class, associations and friendships should be freely formed within the total student body without class distinction. And status should be determined by academic progress and other accomplishments, not time served. That would be the democratic way; and it would benefit not only the gifted but also those who for some reason take longer than the usual time to finish an academic program.

Outcomes of "Acceleration"—and "the Cold War"

As already indicated in passing, the evidence is that most able students do not suffer from acceleration. But much evidence goes further. *They actually seem to gain!* When Terman compared those in his gifted group who graduated from high school young (under 15½) with those graduating near the usual age (over 16½) he found that 16% more of the younger group graduated from college, and 19% more took one or more years of graduate work. The young group married over a year younger, and had fewer divorces. And twice as many of the younger group were highly successful vocationally. Results to date of the Fund for the Advancement of Education, and of the Ohio State University investigations, are similar; these last studies show both those entering young and those taking a four year program in three years doing better in college, *and* having more successful careers after, than matched cases of the same ability but entering at the usual time or taking the usual four years for a degree.[4] And in most of the investigations to date the accelerates were not carefully selected as capable of rapid progress, or given guidance in acceleration: also, they moved forward mostly by these least satisfactory methods—grade-skipping or lengthened school year. There is every reason to suppose that better selection, guidance, and methods would bring yet better outcomes. For instance, 25% more of a group of bright freshmen selected as capable of acceleration and given guidance in so doing, finally obtained a degree than a controlled group not accelerating. A more effective education in *less* time—how can this be explained? Presumably because more rapid progress is normal for the gifted, and the beginning of adult career (and marriage and responsible citizenship) without long delays is biologically sound. There are of course certain obvious advantages, as that saved time and funds leave more of both for advanced training. There are also motivational gains of possible pervasive importance. It is in the American tradition, and stimulating to the bright and ambitious youngster, that there be opportunity to get ahead. The lockstep negates all that. You may have heard the jibe that a certain state penitentiary had a major advantage over the neighboring university: you could get out of the pen sooner if you did well. *Wise means for "acceleration" are legitimate opportunities for the able to forge ahead, which encourage a general climate of enterprise and lively effort.*

---

[4] The Fund for the Advancement of Education. Bridging the gap between school and college, 1953. Pressey, S. L. Educational acceleration: Appraisals and basic problems. Ohio State Univer., 1949.

In Résumé and Final Application

Numerous studies of human development thus show that able youngsters should progress in school faster than the lockstep rate of a grade a year, not only because they develop faster but to prevent too long drawnout education from delaying productive careers. Wise methods of acceleration expedite progress without hampering social adjustment. With acceleration, more able students complete advanced training; they get into productive careers earlier. And these careers tend to be more fruitful. In addition, congested schools are relieved to some degree. It seems a not unreasonable estimate that each year there remain in the secondary schools perhaps 300,000 bright youngsters who ought well have been graduated. *More trained men, sooner, at less cost,* to counter the Russian technological threat! Clearly such a program calls for frequent appraisals from the primary years to graduate school, continuing guidance, and informed educational statesmanship. Surely, then, the topic is most appropriate to this conference.

College Admission with Advanced Standing

WILLIAM H. CORNOG

The School and College Study of Admission with Advanced Standing originated in discussions of the faculty of Kenyon College regarding the possibility of revising some of the rules governing requirements for the bachelor's degree, in order to enable very able students to save time and yet not lose the essential values of a liberal arts education. President Chalmers of Kenyon described the plan to friends and associates in schools and colleges and in 1951 a group of twelve institutions formed a Committee on Admission with Advanced Standing. The committee consisted of the administrative heads and representatives of the following institutions: Bowdoin, Brown, Carleton, Haverford, Kenyon, Massachusetts Institute of Technology, Middlebury, Oberlin, Swarthmore, Wabash, Wesleyan, and Williams. At a meeting in the spring of 1952, the college presidents and deans agreed to invite into the Central Committee twelve headmasters, principals, and superintendents and in May, 1952, the full committee of twenty-four met to organize what was thenceforth known as the School and College Study of Admission with Advanced Standing.

The Central Committee agreed that it would limit the study to eleven subject fields on the college freshman level, and in the summer of 1952, we proceeded to organize eleven sub-committees, of four colleges and three secondary school teachers each, in the following subjects: English composition, literature, Latin, French, German, Spanish, history, mathematics, biology, chemistry, and physics. The subject matter committees were given the task of defining in their respective subjects the standard of achievement of intensive courses in secondary schools which could be offered to the ablest high school students and for which the twelve colleges could give partial or full first-year credit toward their bachelor's degrees. In addition the Central Committee established from its own membership a Committee on Individual Development to examine and define the qualities of the able student himself and to advise the other sub-committees concerning elements of development not commonly included in the mere acquisition of knowledge.

To insure close contact with the college faculties at every point in the study, the college representatives on the Central Committee named correspondents in their institutions in each subject field. The sub-committees continually sought the advice of the correspondents and kept them informed of the progress of committee work. Committee assignments alone involved the participation of more than a hundred and thirty school and college teachers and administrators.

As is the case with all experiments, the School and College Study began with a series of hypotheses. We made some conjectures about American education and constructed an apparatus to test them and to search for solutions to the problems derived from our postulates. We assumed 1) the continuity of education in school and college, 2) the virtues of the liberal arts continuum, 3) the mutuality of interest and understanding of school and college teachers of the same subject matter, 4) the immediacy of the need to revise the timetable and better to utilize the time of the ablest students, 5) the possibility of describing desirable revisions and means of better utilization in terms of specific subject matter definitions in eleven college freshmen course areas.

Some interesting and challenging propositions may be derived from these basic assumptions. We have been committed to the testing of such concepts as the following:

1. That able students can and should be given more intensive instruction in secondary schools and have the opportunity to qualify for admission to college at a level higher than freshman entrance in specific subjects in which evidence can be shown of strong preparation, the equivalent of first-year college work.
2. That committees of school and college teachers can define requirement for admission with advanced standing acceptable to our twelve participating institutions and broadly applicable in institutions of similar standards and aims throughout the country.
3. That acceleration of able students out of high school after two years or three is less desirable than enrichment of the high school curriculum and admission to college with advanced standing at the normal college entering age of seventeen or eighteen after high school graduation.
4. That the advancement of American education demands the strengthening of our secondary schools, particularly in those divisions in which the ablest students are enrolled, and that the colleges have an obligation to encourage the secondary schools who strive to establish and maintain high standards of academic achievement.
5. That sound learning is respectable and that academic subjects, the content of a liberal arts education, constitute worthy intellectual

and spiritual nourishment for young minds, if these disciplines are liberally and wisely taught.

6. That, finally, this study reaches beyond parochial considerations of departmental regulations and self-interest and even of college degree requirements that it may offer a challenge to American education truly commensurate with the dynamism of our culture, the wealth of our resources, and the still unawakened powers of our highly endowed youth.

Our committees recognized from the start the many difficulties and pitfalls involved in drawing blueprints and specifications by which our principles and propositions could be put to the test of action. They produced definitions of standards which present a broad range of topics, in order to give the secondary school teachers freedom for imaginative teaching and adventurous planning. We have avoided standardizing subjects within narrow limits and have encouraged variety in secondary programs for the gifted. We have tried to plan courses which are liberal also in terms of the academic achievement expected. The cumulative extracting of the best practices and highest standards of our twelve colleges could easily result in an unrealistic, "premium" type of advanced credit. Perhaps this has happened in some of our subjects, perhaps we have priced them out of the market, despite the rich endowments of our gifted students. Nevertheless, we shall not allow definitions and examinations to be the sole criteria for the granting of college credit for school courses. Much weight will also be given to other evidence of unusual ability and valued personal qualities, such evidence as will be found in the recommendations of principals, guidance officers, and teachers. It is, finally, important above all, in our opinion, to encourage and challenge the secondary schools to exercise freedom and a bold use of imagination in their teaching and planning, and to avoid as the plague "cram" courses for the bright student.

In June, 1953, the subject matter committees and the Committee on Individual Development submitted their final reports, which were published and circulated to the twelve college faculties for their vote. In the autumn of 1953, the faculties held discussions of the recommendations of the study and, with no college dissenting, voted approval of the experimental plan to consider for admission with advanced credit able students who had received instruction at the level defined by the committees and had met the standard set by the advanced credit examinations to be administered in the spring of 1954.

In the fall of 1953 also, schools associated with the study indicated their readiness and intention to present candidates for examination and advanced credit in accordance with the conditions of the experiment outlined in the Central Committee's published *Reports of Committees, 1952-1953*. In 1953-1954 about 550 students in eighteen schools were enrolled in courses preparatory to application for advanced credit. In May, 1954, these students were examined in one, two, and in a few instances three and even four subjects. The total number of examination papers was about 1000. In the first year of the experiment, the number of examinations which we have been able to administer has necessarily been limited. This year we have expanded the number of experimental schools to forty-two.

No one can say at this point how many gifted students in how many secondary schools over the country will eventually be offered the opportunity and challenge of advanced courses. We know only that there are probably thousands of very able students who should be given more intensive instruction in our schools. We know that many secondary school teachers and principals are eager to do more than they are doing for their gifted students. We also know that education of the gifted as they should be educated costs money. We shall need, here and there, some allotments of extra teacher time in those high schools which undertake intensive preparation. Many of us already receive extra teacher time for remedial work. Certainly education of the handicapped, of the retarded and the slow learner, costs money and money which we are all glad to spend. We of this study believe in this type of expenditure; we further believe that the provision of such special services for the handicapped is not to be regarded as merely the discharge of humanitarian obligations, nor even to be justified to taxpayers by the arguments of sentiment. The handicapped are members of this society of free men and should receive these services as a birthright, and for these services this society may not take undue credit, as for some singular or added grace. By the same token, we hold that it is no less democratic to provide special educational services for the gifted. This provision is also *their* birthright, for democracy has the responsibility to afford opportunity for full personal development to all its citizens, and to each of them in ways and degrees commensurate with the person's endowment and his aspiration. If this is truly the extent of our society's commitment in education, we face the task of discovering to boards of trustees and boards of education how far short we are falling in meeting that commitment, and what necessary and moderately expensive steps we must take to give our gifted students as full a measure of their educational birthright as we give to their less endowed, and no more than equally deserving fellows.

The advanced examinations which our study administered last May were given not only to the advanced credit candidates in our eighteen secondary schools but also to college control groups who had had comparable instruction in our selected subjects in the twelve colleges. These examinations have been read and scored, and while correlations of secondary school and college performance must be based on somewhat meager numbers it is a fair general conclusion that the secondary school candidates made very respectable showings. This is the more remarkable since it was necessary in many of our schools, in order to get the experiment under way, to make rather hurried revisions and to reach some compromise solutions. The schools are confident that with a few more years' experience a much more satisfactory job can be done. No information is yet available regarding the numbers of our school candidates who have received advanced credit in colleges this fall. It is perhaps significant that many of our schools report that their candidates will probably not seek advanced credit, but are content with the satisfactions of their enriched secondary school experience. This fact recalls the statement made in the report of our Committee on Individual Development:

"The Committee wholeheartedly supports the values of a liberal education and would caution school advisers not to let these opportunities for advanced work become vocational in primary intent. The gifted student should focus on the academic challenge and the possibilities for personal enrichment, and not be urged to think and aim for possible future professional fields too early in his development. We recognize the fact that gifted students are apt to feel a sense of urgency to get on with professional study, and may therefore look upon a plan for advanced credit as merely a time-saving, corner-cutting device toward that end. We have been aware of a tendency to regard both college preparatory courses in school and liberal arts courses in college as mere stepping stones of credit to the happy shore of a professional course of study and a career. The saving of time in school and college, though often important from the point of view of family finances and also by reason of the extraordinary pace of development of an individual student, is not the essence of this program for advanced credit. The essence of what we seek may rather be expressed in terms of compressing the school and college experience of the very able in such a way as to make its impact and meaning continuously more challenging and rewarding, and to awaken and constantly to stimulate the reason, imagination, ethical sensibilities, and capacities for moral commitment and action in our highly endowed youth.

"We deplore but accept the necessity of measuring scholastic progress by credit counts, but we do not accept the mere acquisition of credit as the aim of our plan. We do, however, hold that credit demonstrably earned should be awarded and not be withheld because of scruples of academic bookkeeping. Our chief concern is that the college credit earned in a secondary school course represents an experience in personal growth as well as a mature grasp of the subject itself."

It should, finally, be emphasized that our examinations are but part of the attempt to define courses and standards. The central purpose of this study is to promote and develop courses and teaching in the schools. To this end, we have extended the responsibilities of our examining committees to visiting schools and to the holding of evaluating sessions of school and college teachers. One of the major affirmations of this study has been that almost incalculable good has accrued from the prolonged and frequent association of our school and college teachers. In late June of this year we held three conferences of our teachers, one in chemistry at Kenyon College, one in biology at Wabash, and one in history at Williams. The enthusiasm expressed by the teachers who participated lends support to the final statement of our committee on Individual Development:

"We are confident that school and college teachers who undertake to share the responsibility and delight of educating our gifted students as they should be educated will not only contribute to the founding of strong and mutually beneficial school and college relations, but will help to establish a standard of education in which the nation can take pride."

In summary, The School and College Study plan will, we hope, afford one means of achieving a balance of opportunity in American education by making a more nearly equitable provision for a comparatively neglected minority in our schools, our ablest students. It will also, we hope, help to place a new emphasis on certain neglected or unfashionable necessities in education. Among these are: the necessity of patient effort in learning, the necessity of mastery of one's subject in teaching, the necessity of addressing ourselves to the study of worthy things, the necessity of producing thoughtful men. It may be that we can produce thoughtful men by a curriculum that does not require our students to think, or that conditions them merely to agile adjustments to the trivialities of daily living. The burden of proof of this possibility lies upon those who maintain that the proper study of mankind is how to live by bread alone, in personal security and unassailable eupepsia. I am unpersuaded that the majority of our students are fit only to be taught to earn a living, salute the flag, make orderly exits from burning buildings, and come in out of the rain. To base education in any large scope on the assumption that common men cannot be touched by greatness or uplifted by genius is to conspire with degradation and embrace despair. What is required to bring men into touch with greatness is, of course, superb teaching. For men must be worthy of the things they teach as well as of the things they learn. The School and College Study would, finally, express the hope that education in America can re-dedicate itself to the making of master teachers who can make us thoughtful men.

In March, 1884, in his twenty-first year, a young man who aspired to be a teacher wrote in his diary, "Out of the shabby, squalid, starving life if I come with scholarship about me and in me and on me, these are the things I should long to do." He meant, to lecture on Wordsworth, Emerson, Shakespeare, Dante, Goethe. The young man joined the faculty of Central High of Philadelphia in September, 1884, and for twenty-three years was a master teacher. He taught Shakespeare, and wrote the definitive ten-volume life of Benjamin Franklin. He delivered the oration at the dedication of Franklin's statue in Paris. When news of his untimely death reached Stratford-on-Avon, where he had spent many of his holidays, the local schools closed for the day. His influence upon a generation of men has endured with them, and the mark he made upon the school is ineradicable. His name was Albert Henry Smyth, and he left a legend and a legacy. I suppose the moral of the story is this: He was a graduate only of the Central High School; he never went to college. He was the son of an unsuccessful carpet salesman, but much was expected of him. At the High School, he had to study, was given copious opportunity to study, high and serious things, under teachers who were worthy of the noblest calling. We still have such teachers today. May their tribe flourish and increase!

Special Treatment for Abler Students and Its Relation to National Manpower

---

DAEL WOLFLE

In the year 1820 the professorship of astronomy at Trinity College, which carried with it the title of Royal Astronomer of Ireland, fell vacant. After due deliberation, the electors chose William Rowan Hamilton to this distinguished post, and to the research opportunities it provided. The appointment was a notably successful one, for Hamilton served with distinction until his death and came generally to be acclaimed as one of the two or three greatest mathematicians of the English speaking countries.

The astonishing feature of Hamilton's appointment was his age and status: he was 21 years old and still an undergraduate student at Trinity College. The great boldness of suddenly transforming an undergraduate into a professor marks this as one of the outstanding examples of the early recognition and reward of high ability. But there have been many other, albeit less dramatic, examples. Newspapers and magazines last summer told the story of a coed who earned her bachelor's degree at the University of Illinois in a single year. Students who finish college in less than four years have been known to all of us. Lehman has piled evidence on evidence of the early age at which many great accomplishments in science, literature, and the arts have been made. Professor Pressey's excellent work at the Ohio State University constitutes a model of how to bring a larger number of able students into productive careers at a younger than usual age.

There is a common element in Professor Pressey's work, in the quickly-earned degree of the Illinois coed, in Hamilton's youthful attainment of professorial status: in each case a college faculty recognized and nurtured the outstanding ability of a young man or woman whose interests, earlier attainments, background, and financial resources had brought him into a college in which his ability could come under the observation of interested faculty members. It is important to identify such students and to provide them with opportunities for the full development of their abilities. But we must also remember that there are other highly capable young men and women who do not go to college and who do not come under the eye of an interested faculty member. Because they lack the money or more frequently because they lack academic interest they go to work instead of to college. They constitute a portion of the nation's intellectual resources which we are not now utilizing as effectively as we might.

Before taking up this group in greater detail, let me give a basis for being concerned about the optimal utilization of the nation's intellectual resources. The usual justification consists of a comparison between anticipated supply and anticipated demand for scientists, engineers, doctors, school teachers, or professional men and women in other fields. These supply-demand comparisons are based upon rather crude analyses of the manpower market—crude because our techniques for estimating manpower demand are considerably more primitive than are our techniques for estimating a commodity demand. They result in such estimates as that in 1960 the nation will be short 40,000 doctors, will need so and so many more engineers or school teachers than are in sight, or will have an adequate supply of foresters or lawyers. Despite their limitations, such analyses have value; for a few years into the future they can be of aid in the counseling of students and they provide guide lines for the planning of educational programs, personnel policies, and manpower utilization programs.

But the supply and demand analysis is not the only justification for feeling a serious concern over the problem of how we utilize our intellectual resources:

"A society which permits a significant portion of its members to work at levels below their capabilities is failing to achieve its full potential strength. The ability of a society to progress, the ability to better the goals for which it strives, and the skill and wisdom with which it meets its challenges are likely to be the decisive factors in determining its fate . . . the wisdom of the society is the wisdom of its members. Thus a society can attain its full potential only when each of its members is enabled to contribute as fully as his individual abilities permit.

"Judged by this standard the United States has failed to reach the strength which it might attain, for it wastes the abilities of many of its most capable sons and daughters and thereby loses the contributions they might have made. Our potential is greater than our achievement; progress in science and the arts might go on at a faster pace than it does; the moral strength of the country might be greater than it is.

"This standard is fundamentally different from the supply and demand concept . . . Instead of asking *How many jobs are there for scientists?* it asks *How much scientific talent is not being utilized?* Instead of asking *How many additional social scientists and humanists* *could be employed?* it asks *How much important and useful knowledge about humanity which we might learn are we failing to learn?*"[1]

Several kinds of information are required to answer the question of how much talent is being wasted in the United States. We need to know what happens to bright youngsters who go to college, what happens to equally bright ones who stay out of college, and how many there are in each group. A full examination of these questions would require agreement upon what we mean by good utilization of talent and agreement upon the levels and methods of measuring the ability we are talking about. I will take up these points later in a published account, but let me skip over them here and give only the highlights of a recently completed study.

In cooperation with the school and college officials of Illinois, Minnesota, and Rochester, New York, my colleagues and I in the Commission on Human Resources and Advanced Training studied a group of superior high school graduates of some 20 years ago in order to find out what they were doing and how much they were earning in 1954. We analyzed the differences in their occupations and incomes in terms of the ability they had shown in high school and in terms of the amount of education they had obtained after graduating from high school. The 8,000 or more subjects for whom we had fairly full records came from large and small high schools located in urban, suburban, and rural areas in three regions of the United States.

Ability, whether measured by score on a standardized test or by rank in high school graduating class, was related to adult occupation. Those who in the 1930s had been far above average as high school students were in 1954 more likely to be engaged in professional occupations than were those who were only slightly superior to the average high school graduate. Moreover, the occupational differences remained when we corrected for differences in amount of post-high school education. Higher-level positions were associated with higher ability, among those who had gone to technical schools, among those who had entered college but had dropped out before graduation, and among those who had received college degrees.

Similar differences in earnings were found. Those who had finished high school in the top 20 per cent of their classes received higher incomes in 1954 than did those who finished closer to the high school average. And that also was true within each of several educationally homogeneous sub-groups.

---

[1] Wolfle, Dael. ***America's resources of specialized talent.*** New York: Harper and Brothers, 1954. P. 137.

Because family background might well have contaminated the analysis, we took a look at what had happened to the sons of professional men, the sons of laborers, the sons of farmers, and the sons of men in other occupational groups. In general, family background showed a small influence on adult earnings and positions. For example, the sons of professional men were receiving higher incomes than were the sons of laborers, whether the comparison was made among college graduates, college dropouts, or those who had never attended college.

The differences in occupations and incomes which were associated with differences in ability, and those which were associated with differences in family background, were much smaller than were the occupational and earning differentials associated with differences in education. At any ability level, as ability had been manifest in high school, and for any background, as background was indicated by the father's occupation, college paid off. The superior high school graduates who had spent some time in college were better off, occupationally and financially, in 1954 than were those who had not, and those who had graduated from college were better off than were those who had entered college but had not remained long enough to earn a degree. Within the restricted range of ability with which we dealt it meant a few hundred dollars a year higher income to be near the top of the high school ability distribution than to be near the middle. It meant a few hundred dollars a year more to have the background of a professional home than to come from a laboring family background. But possession of a college degree meant two to three thousand dollars a year higher income than was received by otherwise comparable classmates who went directly to work after graduating from high school. Moreover—and this point I want to stress—the income differential was greatest for those of greatest ability. If income can be accepted as a rough measure of what a man is worth to society, then it seems that a man of moderately superior qualifications can earn more and contribute more to society with a college education than he can without it; a man of very superior qualifications can earn and contribute much more.

These findings lead back to the distinction I drew earlier between able young men and women who go to college and equally able ones who do not. Both groups are part of the nation's intellectual resources. If we are seriously concerned—as I think we must be—about how the nation uses its intellectual resources, we must interest ourselves in both of these groups, the ones who go to college and the ones who do not. The two groups are similar in that they include people of high ability, but the most effective techniques of capitalizing on their abilities are somewhat different, and herein lies a problem of educational and social planning.

On the one hand we have able young men and women who are academically motivated. Because of family interests, educational success, the aspirations of their fellows, or for other reasons, they go to college and sometimes to graduate or professional school. The primary problem for them is to see that they get the best possible education. The primary danger is that they will be lost in the mass of less capable high school and college classmates, will be offered a less stimulating educational fare than would be best for them, and will, as a result, waste a portion of their time and effort and not develop as far as they might with a more appropriate education. There is a growing recognition of the special needs of these students. The programs described by Dr. Pressey and Dr. Cornog are specifically designed to enable them to make better use of their educational years than would be the case if they followed the route of the average student.

The other group includes those who are bright but are not academically motivated. There are substantial numbers of such persons. At an intelligence level one standard deviation above the population average about one man and woman in four currently graduates from college. At a level two standard deviations above the population average, about one in two graduates from college. The number going through college could be increased by offering more scholarships. But financial help is not the only remedy necessary. Interest is more often lacking than money. Encouragement, guidance, and an elementary and secondary school experience that will challenge their abilities are at least as necessary as are scholarships. To bring about these improvements will, obviously, require changes in attitude on the part of many people and will require action at the local community level. The task is one for the entire educational system, and for the society that supports and benefits from that system.

In conclusion I would like to make one final observation. It is this: such evidence as we have indicates that it is worthwhile to give greatest attention to those of highest ability. With limited resources, I would concentrate on the top one or two or five per cent; with more generous resources I would extend the effort over the top five or ten or twenty per cent. There are tens of thousands of young men and women who are or could be above average college students; there are quite a few with very superior ability; there are precious few William Hamiltons. The widely ramifying contributions which the very ablest can make are so important that every reasonable effort should be made to identify and educate all of them. The nation needs the scientific progress, technological achievement, and moral strength that can come from the minds of its ablest sons and daughters. Those of us who can help to identify the youngsters of highest potential, who can point the way to overcoming the obstacles which now keep some of this talent from coming to full development, and who can help to develop the kinds of elementary, secondary, and collegiate education which is best adapted to the training of highly able youth share much of the responsibility for determining the future progress of society.

DISCUSSION

PARTICIPANTS

RALPH BERDIE, HERBERT S. CONRAD, EDWARD E. CURETON,
WALTER N. DUROST, DAEL WOLFLE

DR. CONRAD: Are these people whom Dr. Wolfle was speaking about those who graduated from high school twenty years ago?

DR. WOLFLE: They graduated between 1920 and 1938.

DR. CONRAD: Do you generalize from that generation to the present one?

DR. WOLFLE: In general principles, yes. In terms of amounts and details, no. For example, the earning differential would change as you studied a large fraction of the population or studied graduates of a different time period.

Appendix

Participants—1954 Invitational Conference on Testing Problems

ADKINS, Dorothy C., University of North Carolina

AFFLERBACH, Janet, Professional Examination Service

AHMANN, J. Stanley, Cornell University

ALMAN, John E., Boston University

ALLEN, Kathryn M., Department of Education, Schenectady

ALLEN, Margaret E., Board of Education, Portland, Maine

ALLISON, Roger, Educational Testing Service

ALT, Pauline M., Teachers College of Connecticut

ANDERHALTER, Oliver F., St. Louis University

ANDERSON, Rosa G., Psychological Corporation

ANDERSON, Roy N., North Carolina State College

ANDERSON, T. W., Columbia University

ANDERSON, Mrs. T. W., New York City

ANGELL, George W., Jr., Educational Testing Service

ANGOFF, William H., Educational Testing Service

ANTHONY, William, Maryland State Department of Education

APPEL, Valentine, Greenwich, Connecticut

ARMSTRONG, Fred, Lehigh University

ARONOW, Miriam S., New York City Board of Education

ARSENIAN, Seth, Springfield College, Massachusetts

BALLISTY, Al, Personnel Department, Philadelphia

BANNON, Charles J., Crosby High School, Waterbury, Connecticut

BARDACK, Herbert D., New York State Department of Civil Service

BARNES, Paul J., World Book Company

BARTELME, Phyllis F., Blythedale, New York City

BARTNICK, Robert, Educational Testing Service

BATTISON, Mrs. C., Remedial Education Center, Washington, D. C.

BAYROFF, A. G., Personnel Research Branch, AGO

BEDARD, Joseph A., Public Schools, New Britain, Connecticut

BEMENT, Dorothy M., Northampton School for Girls

BENDA, Harold, New Jersey State Department of Education

BENNETT, George K., Psychological Corporation

BENNETT, Ralph, New York City

BENSON, Arthur L., Educational Testing Service

BENT, Alma, State Teachers College, New Paltz, New York

BENTZ, Jon, Sears Roebuck & Company

BERDIE, Ralph F., University of Minnesota

BERG, Joel, University of Connecticut

BERGER, Bernard, Municipal Civil Service Commission, New York City

BERGESEN, B. E., Personnel Press, Incorporated

BLAUL, R. Elizabeth, Highland Park High School, Illinois

BLOOM, Benjamin, University of Chicago

BOASI, Veronica, Department of Personnel, New York City

BOGAR, Jack, Board of Education, Richmond, Virginia

BOGOSIAN, Dorothy, Queens College

BOLDT, R. F., Educational Testing Service

BOLLENBACHER, Joan, Cincinnati Public Schools

BONNER, Hubert, Columbia University

BRACE, Susan E., Archdiocesan Vocational Service, New York City

BRANDT, Hyman, American Occupational Therapy Association

BRANSFORD, Thomas L., New York State Department of Civil Service

BRAY, Douglas W., Columbia University
BRIDGMAN, Donald S., American Telephone and Telegraph Company
BRISTOW, William H., Bureau of Curriculum Research, New York City
BRODERICK, J., Lawrence, YMCA, New York City
BROGDEN, Hubert E., Personnel Research Branch, AGO
BROLYER, Cecil R., New York State Department of Civil Service
BROOKS, Richard B., College of William and Mary
BROWN, Frederick S., Great Neck Public Schools, New York
BRYAN, Ned, Rutgers University
BRYAN, Miriam M., Silver Burdett Company
BUCKTON, LaVerne, Brooklyn College
BUEL, William D., Temple University
BURDOCK, Eugene, Carnegie Corporation of New York
BURKE, James M., Darien Public Schools, Connecticut
BURKE, Paul J., Bell Telephone Company
BURNHAM, Paul S., Yale University
BUROS, Oscar K., Rutgers University
BYRNE, Lois A., Temple University
BYRNE, Richard Hill, University of Maryland
CADWELL, Dorothy, Civil Service Commission, Canada
CAMPBELL, Donald W., Board of Education, Newark, New Jersey
CAPPS, Marian P., South Carolina State College
CARLSON, C. Ray, Air University, Maxwell AFB
CARLSON, Harold S., Upsala College
CARLSON, J. Spencer, University of Oregon
CARROLL, John B., Harvard University
CELLIERS, Peter J., *Pathfinder . . . The Town Journal*
CHAUNCEY, Henry, Educational Testing Service
CHURCHILL, Ruth, Antioch College
CLARK, Willis W., California Test Bureau
COBB, William E., Pennsylvania State University
COCKLIN, John H., Temple University
COFFMAN, William E., Educational Testing Service
COHEN, Philip S., State Teachers College, Montclair, New Jersey
COLEMAN, William, University of Tennessee
CONRAD, Herbert S., U. S. Office of Education
CONWAY, C. B., Department of Education, Victoria, British Columbia
COPELAND, Herman A., Atlantic Refining Company
CORNEHLSEN, John H., Tufts College
CORNOG, William H., Central High School, Philadelphia
COX, Henry M., University of Nebraska
CRANE, Percy F., University of Maine
CRAVEN, Ethel C., Polytechnic Institute of Brooklyn
CRAWFORD, Barbara, Educational Testing Service
CRAWFORD, J. R., University of Maine
CRISSY, W. J. E., Personnel Development, Inc.
CRONBACH, Lee J., University of Illinois
CRUTCHFIELD, Richard S., University of California
CUMMING, Mary, Boston Public Schools
CURETON, Edward E., University of Tennessee
CURETON, Louise W., Board of Examiners in Psychology, Tennessee
CYNAMON, Manuel, Brooklyn College
DAHNKE, Harold, Michigan State College
DAILEY, John T., Bureau of Naval Personnel
DALY, Alice T., New York State Department of Education
DAMRIN, Dora E., Educational Testing Service
DAVIDOFF, M. D., U. S. Civil Service Commission
DAVIDSON, Helen H., City College of New York

Davis, Fred B., Hunter College

Davison, Hugh M., Pennsylvania State University

Dees, Bowen C., National Science Foundation

Degan, James W., Massachusetts Institute of Technology

DeVinney, Leland C., Rockefeller Foundation

Diamond, Lorraine K., Teachers College, Columbia University

Dick, G. W., International Business Machines

Diederich, Paul B., Educational Testing Service

Dion, Robert, California Test Bureau

Dobbin, John E., Educational Testing Service

Doppelt, Jerome E., Psychological Corporation

Dragositz, Anna, Educational Testing Service

Drake, L. E., University of Wisconsin

Druz, William, New Jersey Civil Service Department, Trenton

Duker, Sam., Brooklyn College

Dunn, Frances E., Brown University

Duran, June C., California Test Bureau

Durost, Walter N., Test Service and Advisement Center, New Hampshire

Dwyer, Paul S., University of Michigan

Dyer, Henry S., Educational Testing Service

Ebel, Robert L., State University of Iowa

Edrington, T. C., Department of Defense

Edwards, Robert E., Rensselaer Polytechnic Institute

Engelhart, Max D., Chicago City Junior College

Epstein, Bertram, City College of New York

Estavan, Donald, Educational Testing Service

Ewers, Dorothea, Bloom Township High School, Chicago Heights

Fan, C. T., Educational Testing Service

Farnum, Henry M., Human Engineering Devices

Farr, George C., International Business Machines

Feinberg, Mortimer R., Bernard M. Baruch School and Research Institute of America

Feldt, Leonard S., State University of Iowa

Fendrick, Paul, Western Electric Company

Ferguson, William C., World Book Company

Ferris, F. L., Jr., Educational Testing Service

Fields, Mrs. Carlyle, Croydon Hall, Atlantic Highlands, New Jersey

Fifer, Gordon, Test Research Service, Inc.

Findley, Warren G., Educational Testing Service

Fink, August A., Jr., Columbia University

Finkle, Robert B., Metropolitan Life Insurance Company

Firth, Louise, Educational Testing Service

Flanagan, John C., American Institute for Research

Fletcher, Frank M., Ohio State University

Folger, John, Southern Regional Education Board

Forlano, George, Board of Education, New York City

Forlano, Mrs. George, Board of Education, New York City

Forrester, Gertrude, West Side High School, Newark

Fox, William H., Indiana University

Frederiksen, Norman, Educational Testing Service

Freeman, Paul M., Educational Testing Service

French, Benjamin J., New York State Department of Civil Service

French, John W., Educational Testing Service

Friedenberg, Edgar Z., Brooklyn College

FRIEDMAN, Sidney, Bureau of Naval Personnel
FULTON, Ailsa W., American and Foreign Teachers' Agency
FULTON, Renee J., Board of Education, New York City
FURST, Edward J., University of Michigan
GALLAGHER, Henrietta L., Educational Testing Service
GARDNER, Eric F., Syracuse University
GAWOSKI, Roman S., Yonkers, New York
GERBERICH, J. R., University of Connecticut
GIDDINGS, Frank, Springfield Trade High School, Massachusetts
GLASER, Robert, American Institute for Research
GLASS, Albert A., The Signal School, Fort Monmouth
GODDARD, W. A., International Business Machines
GODSHALK, Fred I., Educational Testing Service
GOLDSTEIN, Leo S., Teachers College, Columbia University
GOODMAN, Joan, World Book Company
GOODMAN, Samuel M., Board of Education, New York City
GORDON, Mary Alice, Macy's
GRIMM, Elaine R., Professional Examination Service
GREEN, Bert F., Massachusetts Institute of Technology
GROVES, Kenneth J., Air University, Maxwell AFB
GUTHRIE, George M., Pennsylvania State University
GULLIKSEN, Harold, Educational Testing Service
HAERTEN, Heinz, Studienstiftung, Bad Godesberg, Germany
HAGEN, Elizabeth, Teachers College, Columbia University
HAGMAN, Elmer R., Board of Education, Greenwich, Connecticut
HALPERN, Joseph, New York State Department of Civil Service
HARMON, Lindsey B., National Research Council
HARTER, R. K., American Telephone and Telegraph Company
HARVEY, Philip R., University of Connecticut
HASTINGS, J. Thomas, University of Illinois
HAYWARD, Priscilla, Educational Testing Service
HEATON, Kenneth L., Richardson Bellows, Henry and Company
HEIL, Louis M., Brooklyn College
HELMICK, John S., Educational Testing Service
HELMSTADTER, G. C., Educational Testing Service
HEMPHILL, John K., Ohio State University
HILL, Walker H., Michigan State College
HILLS, John R., Educational Testing Service
HITTINGER, William F., Pennsylvania State College
HOBERMAN, Solomon, Department of Personnel, New York City
HOLLEY, Clifford S., Personnel Department, Philadelphia
HOLLIS, William H., New York City
HOLLISTER, John S., Educational Testing Service
HOLMES, A. F., Royal Canadian Air Force
HOLMES, Therese, Metropolitan Life Insurance Company
HOROWITZ, Milton W., Queens College
HORTON, Clark W., Dartmouth
HUDDLESTON, Edith, Educational Testing Service
HUNSICKER, Paul, University of Michigan
HUNTER, Genevieve, P., Archdiocesan Vocational Service
HURLBURT, Allan S., North Carolina State Department of Public Instruction
HYLLA, Erich, Hochschule Internationale Paedagogische Forschung, Germany
JANOWSKY, John, Personnel Department, Philadelphia

Jarecke, Walter H., West Virginia University
Jaspen, Nathan, National League for Nursing
Jeffrey, W. E., Barnard College, Columbia University
Johnson, A. Pemberton, Educational Testing Service
Johnson, Lewis W., Personnel Department, Philadelphia
Jordan, Arthur M., University of North Carolina
Kaback, Goldie R., City College of New York
Kalin, Robert, Educational Testing Service
Karon, Bertram P., Educational Testing Service
Keats, J. A., Educational Testing Service
Kidd, John W., Michigan State College
Kiger, Joseph C., American Council on Education
Kimball, Elizabeth G., Educational Testing Service
King, Richard G., College Entrance Examination Board
Kipnis, David, New York University
Kirkpatrick, Forrest H., Bethany College
Kolkebeck, Robert F., Educational Testing Service
Kosmerl, Alice F., Washington, D. C.
Kostick, M. M., State Teachers College, Boston
Krathwohl, David R., University of Illinois
Kubis, Joseph F., Fordham University
Kuder, G. Frederic, Duke Universtiy
Kurtz, Albert K, University of Florida
Kushner, Rose E., City College of New York
Kutcher, Charlotte, American Public Health Association
Langmuir, C. R., Psychological Corporation
Lannholm, G. V., Educational Testing Service
Layton, Wilbur L., University of Minnesota
Leach, Kent W., University of Michigan
Leavy, Sylvia, Queens College
Lebold, William K., Purdue University
Lee, Marilyn C., Science Research Associates
Lennon, Roger T., World Book Company
Lev, Joseph, New York State Department of Civil Service
Levine, Richard, Educational Testing Service
Levy, Charlotte, Chunky Chocolates
Limburg, Charles C., U. S. Air Force
Lindquist, E. F., State University of Iowa
Lohman, Maurice A., New York State Department of Education
Loos, Gordon M., Educational Testing Service
Lorch, Vera, Great Neck, New York
Lord, Frederic, Educational Testing Service
Lord, Shirley, Educational Testing Service
Lorge, Irving, Teachers College, Columbia University
Lorr, Maurice, Veterans Administration
Lund, Kenneth W., Public Schools, Chicago
Lusk, Louis T., Norwalk, Connecticut
Lutz, Orpha M., State Teachers College, Montclair, New Jersey
McArthur, Charles, Harvard University
McCabe, Frank J., Metropolitan Life Insurance, New York City
McCall, W. C., University of South Carolina
McCambridge, Barbara, Educational Testing Service
McCann, Forbes E., Personnel Department, Philadelphia
McCord, Richard B., Personnel Department, Philadelphia
McEwan, Dorothy M., New York State Department of Civil Service
McGill, William J., Massachusetts Institute of Technology

McGinnies, Elliott M., University of Maryland
McIntosh, Vergil M., Air University, Maxwell AFB
McLaughlin, Eugenia, New York State Department of Civil Service
McNamara, W. J., International Business Machines
MacPhail, A. H., Brown University
McQuitty, John V., University of Florida
Macaluso, Charles J., U. S. Naval Examining Center
Machi, Vincent S., Columbia University
Manuel, Herschel T., University of Texas
Marriott, John C., World Book Company
Marston, Helen M., Educational Testing Service
Masia, Bertram B., New York University
Mathews, Chester O., Ohio Weslayan University
Mathewson, Robert H., Division of Teacher Education, New York City
Maxson, Georgia, Educational Testing Service
Medley, Donald M., Division of Teacher Education, New York City
Melville, S. Donald, Educational Testing Service
Merenda, Peter F., U. S. Naval Examining Center
Merry, Robert W., Harvard University
Metz, Elliott, Queens College
Michael, William B., University of Southern California
Michael, S. R., Educational Testing Service
Micheli, Gene, Metropolitan Life Insurance Company
Miles, Matthew B., Teachers College, Columbia University
Mill, Cyril R., Public Schools, Richmond
Millett, Esther, Westover School, Middlebury
Mitchell, Blythe C., World Book Company
Mitzel, Harold E., Division of Teacher Education, New York City
Mollenkopf, William G., Educational Testing Service
Morgan, Antonia Bell, Aptitude Associates Incorporated
Morgan, H. H., Psychological Corporation
Morris, John B., University of Minnesota
Morris, Nancy, Educational Testing Service
Morrison, Alexander W., Polytechnic Institute, Brooklyn
Morrison, J. Cayce, Puerto Rican Study
Morton, Anton S., Educational Testing Service
Mosely, Russell, Wisconsin State Department of Public Instruction
Munger, A. M., Standard Oil Company, New Jersey
Murray, John E., Special Devices Center, ONR
Myers, Robert L., Temple University
Nelson, Kenneth D., Division of Research, Albany
Nelson, M. J., Iowa State Teachers College
Nevin, Margaret, Educational Testing Service
Nill, Kathryn F., Rinehart & Company
Noll, Victor H., Michigan State College
North, Robert D., Educational Records Bureau
Norton, Dee W., Air Force Personnel & Training Research Center, Lackland AFB
Olsen, Marjorie, Educational Testing Service
Orleans, Beatrice S., Bureau of Ships, Navy Department
Orleans, Joseph B., George Washington High School, New York City
Pace, C. Robert, Syracuse University
Palmer, Orville, Educational Testing Service
Pashalian, Siroon, Queens College
Pearson, Richard, Educational Testing Service

Perlman, Mildred, Department of Personnel, New York City

Perloff, Robert, Personnel Research Branch, AGO

Perloff, Evelyn, Board of Education, Maryland

Perry, William D., University of North Carolina

Peterson, Donald A., Life Insurance Agency Management Association

Phelan, Robert F., Prudential Insurance Company

Pierson, George A., Queens College

Pinzka, Charles F., Educational Testing Service

Pitcher, Barbara, Educational Testing Service

Plumlee, L. B., Educational Testing Service

Polin, A. Terrence, U. S. Air Force, Mitchel AFB

Pollack, Norman C., New York State Department of Civil Service

Pressey, Sidney L., Ohio State University

Pressey, Mrs., Sidney L., Ohio State University

Rabinowitz, William, Bank Street College of Education, New York City

Radasch, John, Department of Education, Concord, New Hampshire

Rapparlie, John H., Owens—Illinois

Raskin, Evelyn, Brooklyn College

Raskin, Judith, Queens College

Raymond, Thomas J., Harvard University

Raymond, Vincent R., Newark College of Engineering

Read, Thomas, Peddie School, Hightstown, New Jersey

Regan, James J., Special Devices Center, ONR

Reichler, Melvin, Queens College

Remmers, H. H., Purdue University

Reuter, W. H., Educational Testing Service

Rhule, Warren A., U. S. Armed Forces Institute

Richardson, M. W., Richardson, Bellows, Henry and Company, Incorporated

Richardson, Ruth P. NYSPA, New York City

Ricks, J. H., Jr., Psychological Corporation

Rimalover, Jack K., Educational Testing Service

Robbins, Irving, Queens College

Roca, Pablo, Department of Education, Puerto Rico

Rock, Robert T., Jr., Fordham University

Rosenblatt, Frank, Cornell University

Rosner, Benjamin, Columbia University

Rulon, P. J., Harvard University

Sadacca, Robert, Educational Testing Service

Saunders, David, R., Educational Testing Service

Sait, Eward, Rensselaer Polytechnic Institute

Sawin, E. I., Air Force ROTC Headquarters, Montgomery

Scates, Alice Y., U. S. Office of Education

Scates, Douglas E., American Social Hygiene Association

Schapiro, Harold B., Great Neck, New York

Schrader, William B., Educational Testing Service

Scott, C. Winfield, Vocational Counseling Service, Inc.

Seashore, Harold G., Psychological Corporation

Sebald, Dorothy D., Hunter College

Sebald, J. F., Worthington Corporation

Seibel, Dean W., Harvard University

Sforza, Richard F., New York State Department of Civil Service

Sharp, Catherine G., Educational Testing Service

Shaycoft, Marion F., American Institute for Research

Shimberg, Benjamin, Educational Testing Service

Siegmann, P. T., American Telephone & Telegraph

Silberman, Harry, City College of New York
Simpson, Mrs. Elizabeth A., Illinois Institute of Technology
Smith, Allan B., University of Connecticut
Smiley, Marjorie B., Hunter College
Smit, Jo Anne, Washington, D. C.
Smith, Denzel D., Office of Naval Research
Smith, Louise, Educational Testing Service
Smith, Muriel, Educational Testing Service
Solomon, Robert, Educational Testing Service
Sorenson, Garth, University of California
Souther, Mary T., Tower Hill School, Wilmington
Spaney, Emma, Queens College
Spaulding, Geraldine, Educational Records Bureau
Speer, George S., Illinois Institute of Technology
Spencer, Douglas, New York City
Starch, Daniel, Daniel Starch & Staff
Starch, Mrs. Daniel, Mamaroneck, New York
Stephan, Frederick F., Princeton University
Sternberg, Carl, Queens College
Stewart, Mary, New York University
Stewart, Naomi, Educational Testing Service
Stice, Glen, Educational Testing Service
Stokes, Thomas M., Metropolitan Life Insurance Company
Stone, Paul T., Huntingdon College
Stuart, William A., Educational Testing Service
Stulbaum, Harold, Metropolitan Life Insurance Co.
Sullivan, Daniel S., New York State Department of Civil Service
Sullivan, Richard H., Educational Testing Service
Super, Donald E., Teachers College, Columbia University
Swineford, Frances, Educational Testing Service
Symonds, Percival M., Teachers College, Columbia University
Tasso, Charles A., American Cyanamid
Tatsuoka, Maurice, Harvard University
Taylor, John F., Woodmere Academy
Taylor, Justine N., Educational Testing Service
Templeton, Hugh, New York State Department of Education
Terral, J. E., Educational Testing Service
Thorndike, Robert L., Teachers College, Columbia University
Thurstone, L. L., University of North Carolina
Traxler, Arthur E., Educational Records Bureau
Triggs, Frances, Commission on Diagnostic Reading Tests
Tucker, Anthony C., Department of Defense
Tucker, Ledyard R, Educational Testing Service
Tukey, John W., Princeton University
Tukey, Mrs. John W., Princeton, New Jersey
Turnbull, William W., Educational Testing Service
Twyford, Loran C., Special Devices Center, ONR
Vickery, Verna L., Southeastern Louisiana College
Voss, Harold A., Special Devices Center, ONR
Wadell, Blandena C., World Book Company
Wahlgren, Hardy L., State Teachers College, Geneseo, New York
Walker, Helen M., Teachers College, Columbia University
Wallace, Wimburn, L., Psychclogical Corporation
Walsh, B. Thomas, Personnel Department, Philadelphia
Walsh, John J., Boston College
Walton, Wesley, W., Educational Testing Service

Waltzman, Hal, New York State Department of Civil Service

Watson, Walter S., Cooper Union

Wellck, A. A., University of New Mexico

Welna, Cecilia T., University of Connecticut

Wenzel, Bernice M., Barnard College

Wesman, A. G., Psychological Corporation

White, George L., Silver Burdett Company

Whitla, Dean K., Harvard University

Whitney, Alfred G., Life Insurance Agency Management Association

Wilke, Marguerite M., Board of Education, Greenwich

Wilke, Walter H., New York University

Wilks, S. S., Princeton University

Willard, Richard W., Harvard University

Williams, Robert J., Columbia University

Williams, Roger K., Morgan State College, Baltimore

Wilson, Kenneth M., Princeton University

Wilson, Phyllis C., Queens College

Winans, S. David, New Jersey Department of Education

Wingo, Alfred L., Virginia State Department of Education

Winterbottom, John A., Educational Testing Service

Wittenborn, J. R., Rutgers University

Wolfle, Dael, American Association for the Advancement of Science

Wood, Ray G., Ohio State Department of Education

Wright, Wilbur H., Teachers College, Geneseo, New York

Yogman, Harvey, University of Denver

Zalkind, Sheldon S., Brooklyn College

Zubin, Joseph, Columbia University